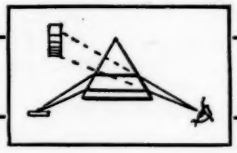
Journal of the OPTICAL SOCIETY *of* AMERICA

VOLUME 26
No. 1



JANUARY
1936

Theory of the Multiple Diffraction Grating . . . . . NORMAN HILBERRY 1

Summation of Different Color Radiations by a Photographic Emulsion . . . . . J. H. WEBB 12

Refraction of Tempered Plate Glass . . . . . JOHN J. HOPFIELD AND GEORGE B. WATKINS 24

Pressure Effect on Sensitivity of Different Photographic Films . . . . . NY TSI-ZÉ AND LÜ TA-YUAN 26

Effect of Oxygen Upon the Photoelectric Thresholds of Metals . . . . . H. C. RENTSCHLER AND D. E. HENRY 30

Simplified Theory of the Image Errors According to L. Seidel . . . . . M. HERZBERGER 35

Theory of Microscope Vision . . . . . M. HERZBERGER 52

Method for Making and Testing Quarter-Wave Plates . . . . . RICHARD M. EMBERSON 63

Instrument for the Measurement of the Graininess of Photographic Materials . . . . . E. M. LOWRY 65

On a Method of Decreasing the Reflection from Nonmetallic Substances . . . . . JOHN STRONG 73

Necrology 75

Published monthly
at Prince and Lemon Sts., Lancaster, Pa.
by the
AMERICAN INSTITUTE OF PHYSICS
Incorporated
11 East 38th Street, New York, N. Y.
FOR AND UNDER THE EDITORIAL SUPERVISION OF
THE OPTICAL SOCIETY OF AMERICA

THE OPTICAL SOCIETY OF AMERICA

OFFICERS OF THE SOCIETY

A. C. HARDY, *President*
Massachusetts Institute of Technology
Cambridge, Massachusetts

R. C. GIBBS, *Vice President*
Department of Physics
Cornell University
Ithaca, New York

L. B. TUCKERMAN, *Secretary*
Bureau of Standards
Washington, D. C.

HENRY F. KURTZ, *Treasurer*
Bausch & Lomb Optical Company
Rochester, New York

The Board of Directors consists of the above officers; the Editor of the Journal; the Past President; and five elected members:

W. B. RAYTON, *Past President*
Bausch & Lomb Optical Company
Rochester, New York

F. K. RICHTMYER, *Editor*
Cornell University
Ithaca, New York

Elected Members

ARTHUR H. COMPTON
Ryerson Physical Laboratory
University of Chicago
Chicago, Illinois

K. S. GIBSON
Bureau of Standards
Washington, D. C.

G. R. HARRISON
Department of Physics
Massachusetts Institute of Technology
Cambridge, Massachusetts

HERBERT E. IVES
463 West Street
New York, New York

CARL W. KEUFFEL
Keuffel & Esser Company
Hoboken, New Jersey

THE OPTICAL SOCIETY OF AMERICA was organized in 1916 *"to increase and diffuse the knowledge of optics in all its branches, pure and applied, to promote the mutual interests of investigators of optical problems, of designers, manufacturers and users of optical instruments and apparatus of all kinds and to encourage cooperation among them."* It aims to serve the interests of all who are engaged in any branch of optics from fundamental research to the manufacture of optical goods.

MEMBERSHIP

Regular Members: "Any person who has contributed materially to the advancement of optics is eligible for regular membership."

Associate Members: "Any person who is interested in optics is eligible for associate membership."

Corporate Members: "Any corporation desirous of encouraging the work of the Society is eligible for corporation membership."

The Society cordially invites to membership all who are interested in any branch of optics, either in research, in instruction, in optical or illuminating engineering, in the manufacture and distribution of optical goods of all kinds or in physiological and medical optics.

ANNUAL DUES

Regular members, $7.50; Associate members, $5.00; Corporate members, $50.00. These dues include subscription to the Society's official journal, the *Journal of the Optical Society of America.*

MEETINGS

The Society holds two meetings per year: An annual meeting in the autumn and a midwinter meeting jointly with the American Physical Society.

For further information concerning the Optical Society of America address the Secretary, Dr. L. B. Tuckerman, Bureau of Standards, Washington, D. C.

ROCHESTER, NEW YORK SECTION OF THE OPTICAL SOCIETY OF AMERICA

C. C. NITCHIE, *President*
H. E. SEEMANN, *Secretary*

L. A. DUBRIDGE, *Vice President*
H. G. KINGSLAKE, *Treasurer*

Councilors: E. C. FRITTS, R. KINGSLAKE, H. F. KURTZ, T. R. WILKINS, C. J. STAUD (*ex-officio*)

Journal of the
OPTICAL SOCIETY *of* AMERICA

VOLUME 26 JANUARY, 1936 NUMBER 1

Theory of the Multiple Diffraction Grating

NORMAN HILBERRY, *New York University, University Heights*

(Received September 18, 1935)

The theory of a multiple diffraction grating is presented. It is shown that the resolving power is that of a simple grating of equivalent ruled aperture; the intensity, that of a simple grating of an equal number of ruled apertures; and the dispersion and resulting distribution of intensity with order, that of one of the elementary gratings of which it is composed. The relative intensities and the distribution of the secondary maxima arising due to the periodicity of spacing of elementary gratings are discussed. The large spectral range, the relatively high intensity and the requirement of little or no preliminary resolution are presented as advantages over interference spectroscopes of equivalent resolving power. The further advantage is presented that the time required for ruling such a grating is so reduced that gratings of two, three or more times the present effective aperture are made readily possible in the same ruling time.

INTRODUCTION

THE trend of spectroscopic work in recent years has been such as to place an ever increasing premium on spectroscopes of high resolving power. In consequence, some consideration as to whether any possible high resolving power instrument had so far escaped discussion seemed to be in order. With this idea in mind certain experiments were carried out in which various component parts of an interferometer were replaced by diffraction gratings. The chief importance of these experiments for the present discussion lay in the fact that they pointed the way directly to the possibility of what might best be called a "multiple grating." For example, an interferometer was adjusted in such fashion that, in effect, the light coming to the observing telescope was that from two similar, coplanar gratings with parallel rulings, but the one displaced laterally with respect to the other. As might have been anticipated, superposed on the spectrum line and parallel to it were the characteristic "double slit" interference fringes. As the effective lateral displacement of the gratings was increased, these "double slit" fringes became narrower in exactly analogous fashion to the true double slit fringes superposed on the single slit diffraction pattern. This experiment is equivalent, of course, to the situation arising whenever it becomes necessary to block out the central portion of a diffraction grating due to errors in ruling. The observed spectrum lines in such a case are those due to a grating of the resolving power of one of the resulting elements crossed by double slit fringes whose separation is determined by the spacing of the two elements. Since, in general, the component gratings are not exactly equal, the actual intensity distribution usually becomes much more complicated.

The obvious extension of these observations

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

was to carry the analogy to its logical conclusion and determine the effect of ruling a grating of gratings. If such a "multiple grating" should actually possess the resolving power which would be suggested by such an analogy, a given number of rulings could be so distributed over a surface as to increase their resulting resolving power considerably and yet retain the advantages inherent in the fundamental ruling separation of the component gratings. At the time this investigation started, one of the factors definitely limiting the ruled aperture obtainable in a grating was the endurance of the diamond ruling points. Another was the length of time requisite for the ruling process with the difficulties attendant upon maintaining invariance of all adjustments throughout this period. Consequently, the multiple grating gave promise of satisfactorily reducing two of the factors most seriously limiting the resolving power obtainable with a ruled grating.

The late Lord Rayleigh was apparently the first to mention the possibility of such a multiple ruling. He pointed out that ruling two apertures and skipping two would be of great value in that the spectrum formed by such a grating would lack the second, sixth, etc., orders and that, consequently, it should be possible to examine the first-order red and the third-order blue without difficulties due to the overlapping of the second order. Following this, Rowland, in his discussion of the effect of shape and arrangement of rulings on the distribution of energy in the spectrum,[1] extended Lord Rayleigh's case to that of a "multiple periodic ruling." As far as the multiple grating is concerned, he confined his attention entirely to the case of a double groove and largely to the specific suggestion of Lord Rayleigh. Of course, Rowland was firmly convinced that the natural breadth of a spectral line was such that the resolving powers then available were more than ample for any future need and his discussion is concerned solely with the effect of the multiple ruling on the distribution of intensity with order. Because of this divergence in objective between Rowland's work and that of the present discussion, it has proved not only more convenient but much more illuminating actually to follow out the analogy suggested and develop the expression for the intensity distribution due to a multiple grating with the standpoint of resolving power in mind rather than to deduce it from Rowland's general analytical expression although the latter course, properly interpreted, is a possible one.

The expression so developed immediately justified the expectation that the resolving power of the multiple grating would be exactly that of a simple grating of the same aperture. Since, however, one must always pay in some fashion for such advantages, the expression also indicated the existence of the set of "secondary" maxima corresponding to primary maxima for a grating with a ruling separation equal to the inter-grating separation of the multiple grating. These secondary maxima were an obvious handicap although it was possible to control within limits both their distribution and their relative intensities.

At this point three lines of investigation were indicated. The first was a detailed examination of the actual intensity distribution curves for various distributions of a given number of apertures. The second was the actual ruling of such a test grating. The third was a careful consideration of the relationship of this multiple grating to the other high resolving power spectroscopes. With the assistance of Messrs. L. K. Toole and J. Leiner of the Gibson Committee of New York City the intensity expressions for several different families of aperture arrangement were computed and plotted for sufficiently small phase intervals to establish completely their entire energy distribution curves. After much grief, and then largely by main force, the author ruled three gratings of 900 effective apertures; one of 900 rulings, one of 450 rulings and one of 225 rulings. These were ruled simultaneously by proper choice of the length of the ruling so that ruling errors were common to all three. As no ruling engine was available, the ruling was done with a micrometer slide and a milling machine. The resulting gratings, as might be expected under the circumstances, were of extremely poor definition but they did show all the salient features to be expected of them.

For the third item, the brevity requisite in a paper of this nature precludes the presentation

[1] H. A. Rowland, Phys. Papers, 525 or Phil. Mag. **35**, 397 (1893); Astronomy and Astrophysics **12**, 129 (1893).

of any satisfactory discussion of the inter-relationships of the various high resolving power instruments and of their points of strength and weakness. This much, however, should be said. Although the Fabry-Perot and Lummer-Gehrcke instruments are in general quite satisfactory, they do require preliminary resolution of the radiations to be examined and even then they are handicapped to some extent by their inherently narrow spectral range. A closely ruled diffraction grating of an aperture such as to give it a resolving power equal to that of the other instruments is free from their handicaps and, moreover, gives a much greater intensity. As the resolving power of a grating for any given angles of incidence and diffraction depends only on the width of the ruled aperture, the obvious procedure would be to rule the grating with a larger grating space. In the simplest case this solution fails in that the diffracted intensity at large angles becomes very small because of changes which this procedure produces in the single aperture factor of the intensity distribution expression. R. W. Wood has overcome this particular difficulty in his echelette gratings by ruling a shaped groove of such form that practically all of the incident intensity is thrown into one of the high orders. Now, however, the grating is subject to the same difficulties as to narrowness of spectral range because of the overlapping of orders as the so-called interference spectroscopes although to a lesser extent. In the Michelson transmission echelon and the reflecting echelon perfected by W. E. Williams, the same logical procedure is carried one step further. The path difference between successive beams is made very great and all the incident intensity is thrown into one or two orders. These interference gratings have all the narrowness of spectral range of the Lummer-Gehrcke and Fabry-Perot instruments and, like them, require preliminary resolution. While they give high intensity they suffer acutely from all the difficulties arising due to the overlapping of orders.

To maintain the primary advantages of the grating it is then necessary to retain, effectively, its close ruling. For high resolving power this entails the ruling of an extremely large number of apertures with the attendant difficulties already mentioned. While recent advances in the technique of using ground and polished diamonds as ruling points have gone a long way toward eliminating the breakdown of the diamond during the ruling process, the difficulties due to the necessity of maintaining invariance of adjustment throughout the length of time requisite for the ruling process still persist. That these difficulties are real is attested by the fact that the number of large aperture gratings being ruled has not increased in proportion to the improvements made in ruling points as should have occurred otherwise. Consequently, it has seemed as if the presentation of the simple theory of the multiple grating might be justified as offering one means by which the advantages of the closely ruled grating of large aperture could in large measure be retained while the time requisite for its ruling could be reduced to one-half or less that required for a simple grating of the same resolving power and fundamental ruling interval.

Theory and Discussion of the Multiple Grating

A consideration of the simple grating establishes three conditions which the multiple grating must fulfill if it is to meet the demands to be made of it. First, it must possess primary maxima whose dispersion and consequent large angle diffracted intensity depend on the elementary ruling distance, secondly, it must have the characteristic subsidiary grating maxima whose intensity ratios with respect to their primaries are not greater than those for a simple grating and finally, the resolving power must depend on the total ruled aperture of the grating in the usual manner.

To derive the expression for the intensity distribution due to a multiple grating it is only necessary to extend the procedure customarily adopted for treating the simple grating. While admittedly not rigorous, the assumption here made that the grating is an array of identical, coplanar, transparent apertures in an opaque screen is amply sufficient to establish the general properties of the instrument. Furthermore, the case of Fraunhofer diffraction customarily realized in the use of a plane grating is assumed throughout.

Under these conditions, the general expression for the intensity distribution due to an array of exactly similar diffracting apertures is mathematically separable into two factors corresponding to the physical fact that the intensity distribution due to each aperture is identical and that the resulting pattern is determined by the superposition of the separate patterns and, therefore, by the phase differences introduced due to the distribution of the apertures. If i and θ are the angles of incidence and diffraction respectively, I_μ' is the single aperture intensity corresponding to the angles i and θ, and d_i is the separation of a point in the ith aperture from the similar point in the zeroth aperture, the general expression may be written as

$$I = I_\mu'(c'^2 + s'^2),$$

where

$$c' = \sum_i \cos \mu d_i; \quad s' = \sum_i \sin \mu d_i \quad \text{and}$$

$$\mu = (2\pi/\lambda)(\sin i + \sin \theta).$$

The simple grating is an array of n equally spaced slits. In this case, d_i assumes the values d, $2d$, ..., $(n-1)d$ and the above summations lead to the customary intensity expression for the simple grating. The multiple grating, however, consists of an array of equally spaced, parallel and identical simple gratings. If d represents, as before, the distance between similar points in successive apertures and b, the distance between similar points in successive gratings, the summations become

$$\begin{aligned} c' = {} & 1 + \cos \mu d + \cos 2\mu d + \cdots \cos (m-1)\mu d \\ & + \cos \mu b + \cos \mu(b+d) + \cos \mu(b+2d) + \cdots \cos \mu[b+(m-1)d] \\ & + \cos 2\mu b + \cos \mu(2b+d) + \cos \mu(2b+2d) + \cdots \\ & \vdots \\ & + \cos (N-1)\mu b + \cos \mu[(N-1)b+d] + \cdots \end{aligned}$$

and

$$\begin{aligned} s' = {} & \sin \mu d + \sin 2\mu d + \cdots \sin (m-1)\mu d \\ & + \sin \mu b + \sin \mu(b+d) + \sin \mu(b+2d) + \cdots \sin \mu[b+(m-1)d] \\ & + \sin 2\mu b + \sin \mu(2b+d) + \sin \mu(2b+2d) + \cdots \\ & \vdots \\ & + \sin (N-1)\mu b + \sin \mu[(N-1)b+d] + \cdots, \end{aligned}$$

where m is the number of apertures in an elementary grating and N is the number of gratings.

If, now, these expressions are simplified and the complex quantities $(c'+is')$ and $(c'-is')$ are set up, the indicated summations can readily be carried out and the product of these quantities gives

$$c'^2 + s'^2 = \frac{\sin^2 m\mu d/2}{\sin^2 \mu d/2} \cdot \frac{\sin^2 N\mu b/2}{\sin^2 \mu b/2}.$$

The intensity distribution for the multiple grating is, then, that of a single elementary grating multiplied by a factor which depends on the number and separation of the elementary gratings. As will be evident later, there is no advantage, at least in its use, in having the inter-grating separation other than an integral multiple of the ruling space and there are certain definite disadvantages in having it otherwise. If, then, the inter-grating spacing, b, be chosen equal to kd, the final intensity distribution expression becomes

$$I = I_\mu' \frac{\sin^2 m\mu d/2}{\sin^2 \mu d/2} \cdot \frac{\sin^2 Nk\mu d/2}{\sin^2 k\mu d/2}.$$

When the phase angle, $\mu d/2$, is any whole multiple of π radians, both denominators in the above expression become zero and the expression assumes its maximum possible value

$$I = I_\mu' m^2 N^2 = I_\mu'(mN)^2.$$

This set of maxima correspond to the primary maxima of a simple grating and the condition determining their positions

$$\mu d/2 = h\pi \quad \text{or} \quad d(\sin i + \sin\theta) = h\lambda$$

is exactly the same as that for one of the elementary gratings; that is, the multiple grating has a set of primary maxima whose distribution in angle depends only on the fundamental ruling space as in the case of a simple grating, thus satisfying the first of the conditions stated for it. It should be pointed out that mN is the actual number of apertures forming the grating and I_μ' is the single aperture term for a simple grating so that the intensity of these maxima is exactly that of a simple grating of the actual number of apertures ruled.

Whenever the phase angle, $Nk\mu d/2$, takes on values given, approximately, by

$$Nk\mu d/2 = Nkh\pi + [(2j+1)/2]\pi \quad (j=1, 2, \cdots)$$

subsidiary maxima arise which might in the present case be called tertiary maxima. The intensity ratios between these maxima and their associated primaries are, approximately,

$$\frac{I_{\text{III}}}{I_{\text{I}}} = \frac{I_{[h+(2j+1)/2Nk]\pi}}{I_{h\pi}} = \frac{1}{(Nm)^2} \times \frac{\sin^2[(2j+1)m/2Nk]}{\sin^2[(2j+1)/2Nk]} \cdot \frac{\sin^2[(2j+1)/2]\pi}{\sin^2[(2j+1)/2N]\pi}.$$

If N is greater than ten or fifteen and if j is confined to values corresponding to tertiary maxima adjacent to a primary maximum, then

$$I_{\text{III}}/I_{\text{I}} = 1/(j+1/2)^2\pi^2,$$

which is exactly the relationship that holds between the intensities of the subsidiary and primary maxima for the simple grating. The primary maxima of the multiple grating, then, possess the same characteristic set of subsidiary maxima with the same intensity relationships as the simple grating, thus satisfying the second condition stated for its validity.

If, now, the Rayleigh criterion be applied to the present case, the first subsidiary minimum following one of the spectral maxima will fall at a phase angle

$$\mu d/2 = h\pi + \pi/Nk$$

and this must correspond to the primary maximum of the other wave-length

$$\mu' d/2 = h\pi.$$

Then $(\mu'-\mu)/\mu' = 1/hNk$ and $\Delta\lambda/\lambda = 1/hNk$. This, however, is the resolving power of a simple grating of Nk ruled apertures and the third condition laid down above is satisfied.

Consequently, ruling m apertures, skipping $(k-m)$ apertures and then repeating the process N times produces a grating with a dispersion and consequent large angle diffracted intensity corresponding to the fundamental ruling space, with primary maxima of intensity corresponding to a simple grating of Nm apertures and with a resolving power corresponding to that of a grating of Nk apertures.

Distribution and Intensity of Secondary Maxima

The price paid for this increased resolving power lies in the set of secondary maxima arising whenever the phase angle $k\mu d/2 = l\pi$, which, in effect, gives rise to a variety of ghost. This spacing is, of course, that for primary maxima from a simple grating of ruling space, kd. As shown above, the resolving power for a given number of ruled apertures increases as the fraction of the total width actually ruled decreases, while the peak intensity remains fixed in value. An increase in resolving power, however, means a decreased width of each diffraction maximum which, in the present case, means a decrease in the total energy lying within it. Since the total energy thrown into the diffraction pattern must always sum up to the same value, the number of diffracting apertures being constant, a decrease in the total energy lying within the primary maxima must be compensated for by an increase in that lying within these secondary, ghost maxima. If both the total number of apertures and the inter-grating separation are to be kept constant, increased resolving power can only be obtained by decreasing the number of apertures in each elementary grating. If it were possible to continue this process indefinitely without changing the single aperture term, the limiting case would again be that of a simple grating, now with a ruling distance, kd, and with

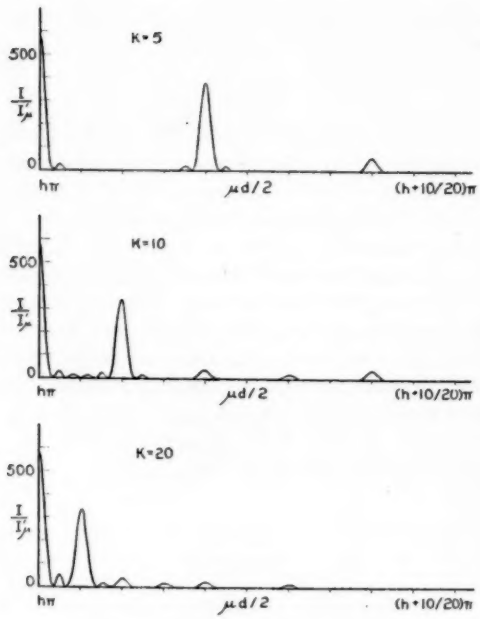


FIG. 1. Effect of inter-grating separation on spectral range $mk=60$; $\sigma=0.4$; $Nm=24$.

primary maxima equal in intensity to those in the first instance and falling at the positions indicated for the secondary maxima of the intermediate cases. However, the effective single aperture term becomes altered radically, certainly for m equal to one and probably also for values of m up to five or even higher. The effect of this change is to give an effectively wider single diffracting aperture which, in turn, gives smaller intensity at the large angles of diffraction. This is, of course, exactly the primary objection already raised to the ruling of gratings of increased ruling interval as furnishing a satisfactory solution of the present problem.

For the intermediate cases, corresponding to the multiple grating, it then becomes essential to examine both the distribution of these secondary maxima and their intensities relative to that of the primary maxima with which they are associated. Obviously, the smaller the inter-grating separation, kd, is made the more widely separated these secondaries will be (Fig. 1). If the spectral range be determined as the difference in wave-length corresponding to that between two primary maxima, the first of which falls upon the first secondary adjacent to the other primary, the conditions can be written

$$\mu' d/2 = h\pi \quad \text{and} \quad \mu d/2 = [h+1/k]\pi.$$

Then $(\mu'-\mu)/\mu' = 1/hk$ or $\Delta\lambda = \lambda/hk$. Thus, it appears that, by proper choice of order and inter-grating separation, the spectral range can be controlled, at least within limits. For example, a k of 100 would give a spectral range of 10A in the third order blue, while a k of ten would give a range of 100A. In this respect the multiple grating differs from the Fabry-Perot or Lummer-Gehrcke instruments. The resolving power of the latter instruments depends on the shape of the intensity distribution curve which is, in turn, determined directly as a function of the reflecting power and not directly as a function of the number of the interfering beams, provided that that number is not too small. For example, in the Fabry-Perot interferometer, an increase of spectral range requires a decrease of etalon distance. This would automatically decrease the resolving power unless accompanied by a corresponding increase in reflecting power, in which case the intensity would be decreased sharply. In the case of the multiple grating, however, the procedure would be to reduce the inter-grating separation by the reduction of the number of apertures per elementary grating, keeping the ratio of k to m constant, and to increase, correspondingly, the number of such gratings. This would produce no change either in the total ruled aperture or in the number of apertures and, consequently, no change either in the resultant resolving power or peak intensity.

The question arises immediately as to the reason for not carrying the above procedure to the limit, that is, for not adopting Lord Rayleigh's double ruling as the elementary grating and thus obtaining the maximum possible spectral range. The primary objection was stated by Rowland in his discussion mentioned above; the fact that in all probability the ruling point cuts the surface in such fashion that the rulings overlap. In such a case the two rulings would not be identical and their pattern would be more or less unpredictable. Rowland was convinced, therefore, that this particular scheme would not work out in practice. Since effects of groove

shape and errors of ruling were so much more vital to the study of the distribution of intensity with order, the ruling of such a grating was apparently never undertaken, although Rowland maintained that it would be a relatively simple matter.

A second objection would lie in the fact that the spectral range obtained by the above procedure would prove to be too great for convenience when the multiple grating is to be used without preliminary resolution. The secondary maxima would be spread throughout the spectrum instead of being confined to the region of their principal maxima and the problem of interpretation due to their overlapping would be increased rather than decreased. The ideal of sharp maxima with relatively large regions of sensibly zero intensity between them should hold for the multiple as well as for the simple grating if the advantage of the simple grating in requiring no preliminary resolution is to be maintained. This means that the secondary maxima should drop sharply in intensity in the neighborhood of their primary maxima and be sensibly of zero intensity elsewhere.

While the obvious method for studying the relative intensities of these secondary maxima would be the straightforward computation and plotting of the desired distributions, the process is tedious and unnecessary, at least for most purposes. Such computations have been carried out for a large number of cases in order to establish the distribution curves completely, but the following procedure is much more satisfactory for present purposes. If the values of $\mu d/2$ which give rise to the secondary maxima are substituted in the general intensity expression, regardless of the value of k, it assumes the form

$$I = I_\mu' N^2 (\sin^2 m\mu d/2)/(\sin^2 \mu d/2).$$

That is, the envelope of all secondary maxima is simply N^2 times the intensity distribution for one of the elementary gratings (Fig. 2). The condition that the secondaries be effectively zero save in the neighborhood of their primary maxima reduces to the condition that the elementary set of rulings be sufficient in number to give the intensity distribution due to a simple grating. Secondary maxima lying within the primary maximum of this simple grating envelope would have, in general, appreciable intensities. Those lying within the first subsidiary maximum of this envelope would have, at the most, an intensity no greater than that of the first subsidiary maximum for the grating as a whole. For long photographic exposures these might give bothersome images but secondaries lying still farther from the primary would drop rapidly to negligible intensity.

Between two primary maxima, the elementary grating distribution curve has $(m-1)$ minima and $(m-2)$ maxima. In this same region there will be $(k-1)$ equally spaced secondary maxima. Thus the number of secondaries lying between a primary and the first minimum of the envelope depends only on the relative values of m and k. As the intensities of these secondaries are determined by their position within the envelope, it is evident that both the number of these ghost maxima which are troublesome and their relative intensities are determined by the ratio of the ruled to the total aperture of the grating. It should be further noted that both the spacing and the relative intensities of these maxima are independent of order in contrast to the Rowland ghosts arising from periodic errors of ruling. Consequently, these ghosts are no more trouble-

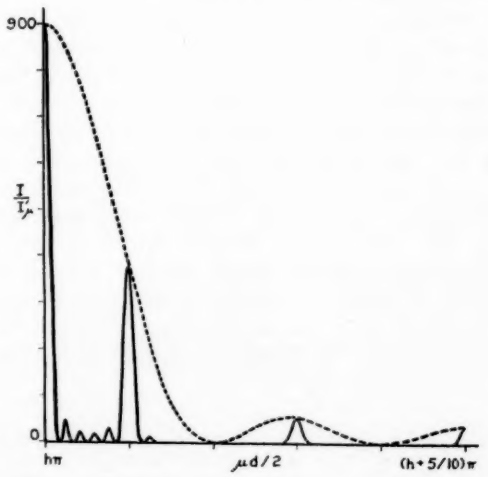


FIG. 2. Full curve—intensity distribution for multiple grating $k=10$; $\sigma=0.5$; $m=5$; $N=6$. Dotted curve—envelope given by N^2 times the intensity distribution due to a single elementary 5-line grating.

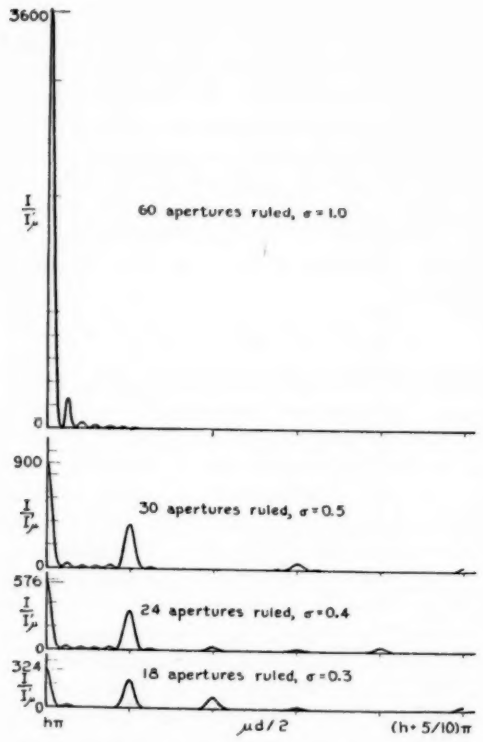


FIG. 3. Intensity distribution curves for a simple grating and multiple gratings of equivalent resolving power $Nk=60$; $k=10$.

some for work in high orders than for low except for the effective reduction of spectral range, already noted, due to the increased dispersion in the higher orders.

Table I gives the approximate ratios, R_1, R_2, etc., of the intensity of successive secondary maxima to that of the primary for various ratios of ruled to total aperture, σ (Fig. 3). As the fraction of the total aperture ruled, σ, gets smaller than 0.25, the number and relative intensity of these secondaries which are troublesome increases rapidly. However, it would become optically and mechanically inconvenient to use a grating of much more than eighteen inches effective ruled aperture. Thus, by considering six inches of ruling as conveniently possible, σ need be only one-third and still smaller values are probably of no practical importance, certainly for the present. Since the intensity of the primary maxima depends on the number of apertures actually ruled, σ should be kept as large as is practicably possible.

For a given value of σ, the problem of confining the secondaries to a given region in the neighborhood of the primary and the magnitude of the spectral range both depend on the choice of k. As the value of k and, consequently, of m is increased the secondaries are confined more closely to the region of the primary. This same effect, however, decreases the spectral range so that the lower limit to this region of confinement is set by the requirements of spectral range which, in turn, fix the upper limit for the value of k. The lower limit for k is fixed by this region of confinement of the troublesome ghosts; that is, any primary, whose ghosts of troublesome intensity could lie within the spectral range being examined, should itself lie within the total spectral region which can be examined at one time, say on a single photographic plate. To satisfy this condition the spectral range must not be greater than the total wave-length range which can be examined at one time divided by twice the number of the last ghost of troublesome intensity. For example, if σ equal 0.5, the last ghost of troublesome intensity would be, probably, the fifth. The spectral range should not be more, then, than one-tenth the wave-length range of a single plate.

For many line spectra and band spectra, these ghosts are likely to be a severe handicap just as the existence of Rowland ghosts has been in the past. If the relatively crude preliminary resolution of filters or a prism of small refracting angle is not barred by experimental requirements, a second mode of procedure is possible which is, of course, equally valid though not as essential in the case of the widely separated single maxima considered above. Suppose that k is made as small as possible. In order to escape Rowland's objection m should probably not be less than ten. To consider a definite case, if σ

TABLE I.

$\sigma=m/k$	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}
0.50	0.406	0.000	0.045	0.000	0.016	0.000	0.008	0.000	0.005	0.000
.40	.573	.055	.021	.040	.000	.016	.004	.003	.007	.000
.33	.685	.171	.000	.043	.030	.000	.014	.011	.000	
.25	.812	.406	.090	.000	.032	.045	.016	.000		

equal one-half, k would be twenty and preliminary resolution to a band fifty Angstroms wide would eliminate all the secondaries within this band in the third order of 3000A giving a spectrum in no way different from that arising if the whole aperture were ruled save that the intensity would be only one-fourth as great.

Since, under the above circumstances, the relative intensities of the secondaries become a matter of minor importance, there is no objection to using still smaller values of m provided experiment should show that such a unit still acted sensibly as a grating in the sense here used and did not suffer from peculiarities due to the unequal size of the last aperture ruled in each group. A reduction of m to a value of five would increase the range in the case above considered to 100A which would be ample for most purposes.

If the inter-grating separation is not an integral multiple of the fundamental ruling interval, the effect is to displace the secondary maxima within the envelope which in turn decreases the peak intensity obtainable in any maximum. Except for orders in which the inter-grating optical path difference is integral, the primary maxima above discussed effectively disappear. To illustrate, if the inter-grating separation were an odd number of halves of the fundamental ruling interval, the denominators would go to zero simultaneously only for even orders and only for such orders would the primary maxima as above defined exist. For the odd orders the optical path difference between successive elementary gratings would be an odd number of half-wave-lengths and the positions of maxima for the elementary gratings would become in consequence minima. Although the primary maxima fail to exist in certain orders due to the fact that no secondary maxima exist for the phase angles corresponding to the peak of the envelope, the fact is of no particular benefit. The effect is to throw two secondaries nearer to the peak of the envelope than the first would normally lie for the case in which the primary maximum exists, and, consequently, give the effect of two somewhat fainter semiprimary maxima in such orders. The ratio of their intensities depends, of course, on the fractional part of the inter-grating separation as measured in units of the fundamental ruling interval. When symmetrically placed about the peak of the envelope the resulting pattern is similar to the double order case for the echelon grating.

While of no particular importance as far as use of the instrument is concerned, the above nonintegral case is of importance from the point of view of the ruling process itself. It should be noted that as the inter-grating separation increases by one ruling interval the diffraction maximum corresponding to the first-order primary shifts toward the central image, becoming the first secondary on that side when the integral inter-grating separation is again attained.

In ruling such a grating two general types of procedure are possible. The first is to utilize the usual ruling engine as it stands, modifying it only in minor details. When it appeared that the life of the diamond was the vital factor, the simple inclusion of a stop which would alternately raise the diamond above the grating surface and then lower it for the desired number of rulings in each case seemed a sufficient modification although it did nothing to shorten the time required for the completion of the grating. It would now appear that the life of the diamond is not the controlling factor and the more effective modification would be the arrangement of a stop which would allow the paul to drop back say ten teeth on the ratchet wheel every eleventh ruling process thus effectively shortening the time required for the ruling and automatically maintaining the integral inter-grating separation.

Another scheme is possible however, and that is to build a doubly periodic machine. As noted above, if the periods were out of the integral relationship by as much as one percent it would only displace the primary maximum from its integral position one percent of its separation from its first secondary in the first order and only three times this amount in the third, which would cause a negligible change in its intensity. A much better adjustment than this could readily be attained in various ways so that such a machine becomes definitely practicable. For low inter-grating separations admittedly requiring some rough preliminary resolution such a machine would be of no particular value. For large values of k, however, the modification of the

present engines might become awkward and a multiply periodic machine would prove of value.

Multiple Gratings by Virtual Images

The possibility of replacing the real elementary gratings above discussed with virtual images should be mentioned. The simplest ways of accomplishing this are, of course, the use of a Fabry-Perot etalon or a Lummer-Gehrcke plate. However, it can also be done by placing a half-silvered plate parallel to a mirror toward one side of which an elementary grating has been ruled or by ruling an elementary grating at one end of a plane parallel plate of glass. Actually, these cases reduce to the two former cases, respectively, with the exception that the interference fringes are now parallel to the spectral line. There are difficulties in practical application, however. Effectively, in both these cases the grating is seen multiply in monochromatic light at a given angle. The successive images may be slid from their observed, to equivalent coplanar positions, the effective inter-grating separation depending on the separation of the two reflecting surfaces and the angle of observation. As this angle changes for each spectral maximum examined, these instruments require either adjustment for each observation or the use of a fixed angle of observation and a variable angle of incidence.

If, however, the whole surface of the mirror or one entire side of the plane parallel plate be ruled and the central zero order beam be multiply reflected, the multiple images are spaced on the equivalent plane at fixed separations independent of the wave-length of the light examined, and thus requiring but one adjustment. Care must be taken that the diffracted beams to be used leave at angles near enough to the normal so that multiple reflections of the diffracted beams may be neglected. While the effective inter-grating separation is determined by the angle of incidence and the optical separation of the grating and the reflecting surface, the effective ratio of ruled to total aperture is determined for a given angle of incidence and effective separation by the width of the incident beam.

Before proceeding further it should be observed that in this case also the intensity of successive images decreases. This decrease arises from two factors, the fraction of the energy removed from the zero order beam by diffraction at each incidence and the fractional loss of energy at each reflection. In the case of the ruled parallel plate this latter factor would, of course, be reduced effectively to zero if the angle of incidence were kept within the critical angle, leaving only the loss upon diffraction and that due to absorption in the plate. In any case the amplitude of the beams apparently arising from successive virtual images is successively reduced so that

$$c' = 1+\cos \mu d+\cos 2\mu d+\cdots$$
$$+f^{1/2}[\cos \mu b+\cos \mu(b+d)+\cdots]$$
$$+f^{2/2}[\cos 2\mu b+\cos \mu(2b+d)+\cdots]+\cdots.$$

The summation for s' is similarly modified and when both are simplified, thrown into complex form as before, summed in the same manner and reexpressed in trigonometrical terms, the result is that

$$I = I_\mu' \frac{\sin^2 m\mu d/2}{\sin^2 \mu d/2} \frac{[f^{N/2}-1]^2-f^{N/2}\sin^2 N\mu b/2}{[f^{1/2}-1]^2-f^{1/2}\sin^2 \mu b/2}.$$

This expression assumes the light to be incident directly upon the grating the first time with an intensity, I_0, and to have an intensity fI_0, f^2I_0, etc., at the second, third, etc., incidences. As mentioned above, no account has been taken of interference due to multiple reflection of the diffracted beams. If the angle of incidence and the separation of the reflecting surface and the grating are such that N is large, $f^{N/2}$ is effectively zero and

$$I = I_\mu' \frac{\sin^2 m\mu d/2}{\sin^2 \mu d/2} \frac{1}{[f^{1/2}-1]^2-f^{1/2}\sin^2 \mu b/2}.$$

The second factor is the intensity distribution expression for Fabry-Perot fringes so that the characteristic subsidiary grating maxima have disappeared and now the pattern is that of Fabry-Perot secondary maxima lying under the elementary grating envelope. Since the resolving power obtainable with these maxima does not increase directly with the number of virtual images as it did with increased number of real

elementary gratings, increased resolving power can only be obtained by decreasing the distance between successive secondary maxima. This, in turn, brings the whole problem effectively back to the use of the Fabry-Perot etalon or the Lummer-Gehrcke plate. While this method of using virtual images is a possible one, its advantages are probably too few to justify its use as compared with the other interference spectroscopes.

Conclusion

While the use of a virtual multiple grating is possible, it appears to offer little, if any, advantage over the regular interference spectroscopes. The same is not true of the ruled multiple grating, however. It occupies a middle ground between the echelette and echelon gratings on the one hand and the simple, closely ruled grating of large aperture on the other. It has the very definite advantages of the latter instrument tempered by the one handicap of a set of characteristic ghosts. Its spectral range, including these ghosts, can be made large enough however, so that with the roughest of preliminary resolution, it can be used even for certain problems in band spectrum analysis requiring high resolving power. It does not rely upon groove form for operation at large diffracting angles and is, therefore, as useful throughout the entire range of angles customarily used as is any ordinary closely ruled grating. Without preliminary resolution it confines the light energy thrown into ghosts to a few secondary maxima within a controllable region in the neighborhood of the primary maxima, thus reducing the task of disentangling the patterns due to overlapping orders.

Finally, without tampering with the fundamental interval of the ruling engine itself, it makes possible the ruling of gratings of any desired effective aperture in lengths of time which are within reasonable limits.

The Summation of Different Color Radiations by a Photographic Emulsion*†

J. H. WEBB, *Eastman Kodak Company*

(Received November 2, 1935)

The problem of determining the manner in which a photographic emulsion sums different color radiations divides itself naturally into two parts depending upon whether the different radiations are applied simultaneously or successively. The first part of this problem was thoroughly investigated by A. Van Kreveld who announced a beautifully simple theoretical law for the manner in which an emulsion adds different colors applied simultaneously. In the first part of the present paper results are presented in verification of Van Kreveld's summation law. The second, and main, part of this paper is concerned with the addition of colors applied successively. A special technique had to be developed for the execution of these experiments and for the reduction of the data. The method of making the exposures consisted of crossing intensity scale curves, and the comparison of the separate and combination color exposures was carried out by graphical means. The results show that an emulsion adds radiations of different color (applied simultaneously or successively) according to the same rule that it adds radiations of identical quality. Van Kreveld's summation law was tested for the case of successive addition and was shown to hold in a restricted form here also. A brief discussion of the theoretical significance of the results is included.

INTRODUCTION

THE problem of determining the manner in which a photographic emulsion sums radiations of different color divides itself naturally into two parts according to whether the different radiations are applied (1) simultaneously, or (2) successively. In both cases it should be possible by a comparison of the effects of the separate color radiations with the effect of a mixture of the same radiations to determine the manner in which an emulsion sums different color radiations.

The first of these problems, namely, that in which the colors are applied simultaneously, was studied first by F. C. Toy[1] who obtained important qualitative results. More recently, this problem has been very completely worked out through an investigation by A. Van Kreveld.[2] From an extensive experimental study of the comparison of separate color exposures with mixture color exposures, Van Kreveld was able to give a quantitative law for the manner in which a photographic emulsion adds different color radiations when applied simultaneously. The simplicity of this law, as well as its rigorous agreement with experiment, goes to make it peculiarly attractive in the field of photographic theory. According to this law, a photographic emulsion adds *correctly* different color radiations, applied simultaneously, in the same sense that an emulsion adds different radiations of identical quality, applied simultaneously.

In the first part of this paper some results will be presented which corroborate the findings of Van Kreveld on the addition of colors applied simultaneously. Since these results are merely confirmatory, they will be given very briefly. The chief purpose of this paper is to give the results of an investigation dealing with the manner in which an emulsion adds different color radiations applied *successively*, and the main portion of the paper will be devoted to this phase of the subject.

Before presenting the results on the addition of colors applied simultaneously, it is thought appropriate to state, and discuss briefly, the Van Kreveld addition law as announced for this case. This will give a logical mode of approach to the experimental results presented here, and at the same time will furnish the basis for comparison of the experimental results with theoretical predictions.

ADDITION OF DIFFERENT COLORS APPLIED SIMULTANEOUSLY; THE SUMMATION LAW

The summation law as given by Van Kreveld for the case of three colors may be written in the form:

* Communication No. 574 from the Kodak Research Laboratories.

† Presented at the annual meeting of the Optical Society of America held in Philadelphia, October 24–26, 1935.

[1] F. C. Toy, Proc. Roy. Soc. **A100**, 109 (1922).

[2] A. Van Kreveld, Zeits. f. Wiss. Phot. Photophysik u. Photochemie **32**, 222 (1934).

$$\frac{R\cdot E_{r+g+b}}{E_r}+\frac{G\cdot E_{r+g+b}}{E_g}+\frac{B\cdot E_{r+g+b}}{E_b}=1 \quad (1)$$

in which E_{r+g+b}, represents the energy of the mixture of colors, and E_r, E_g, and E_b, the energy of the separate colors (monochromatic or otherwise) required to produce a given *density* in the *same exposure time*. The coefficients R, G, and B, represent the fractional ratios of red, green, and blue radiations, respectively, contained in the mixture, and they satisfy the relationship:

$$R+G+B=1.$$

Since all the quantities of Eq. (1) apply only to exposures for which the duration of exposure is the same, the time factor can be eliminated from each term, and the equation can be written in terms of intensities alone, as:

$$\frac{R\cdot I_{r+g+b}}{I_r}+\frac{G\cdot I_{r+g+b}}{I_g}+\frac{B\cdot I_{r+g+b}}{I_b}=1. \quad (2)$$

In order to test the law as expressed in Eq. (2), it would be necessary to measure the intensity of each separate color, as well as the intensity of the mixture of colors, in absolute energy units. The law can be expressed, however, in a slightly altered form and the experiments carried out in a special way, which permits its experimental verification without necessitating the carrying out of these energy measurements. Such a reduced form of expression was used by Van Kreveld in his original experimental verification of the law. Proceeding with this simplification, divide Eq. (2) by the intensity, I_{r+g+b}, which gives the equation:

$$R/I_r+G/I_g+B/I_b=1/I_{r+g+b}. \quad (3)$$

Further, Eq. (3) can be written in the form:

$$1/I_r'+1/I_g'+1/I_b'=1/I_{r+g+b}, \quad (4)$$

in which, I_r', I_g', and I_b', represent intensities, which are not in absolute units but in smaller units than I_r, I_g, and I_b, in the respective ratios R, G and B.

Eq. (4) is in the desired form for testing experimentally. In order to see that (4) is a particularly convenient form, let it be supposed that intensity scale H. and D. curves have been made to the red, green and blue sources separately, and also to these three identical sources simultaneously. Further, let it be supposed that all of these curves have been drawn to the same abscissa scale, i.e., as though, for each curve, the corresponding exposure steps had the same value. This is seen to be equivalent to stretching the abscissa scale for the red, green and blue radiation curves in the respective ratios $1/R$, $1/B$, and $1/G$ or, what amounts to the same thing, of plotting these curves in units which are diminished in the ratios R, G, and B, respectively, as suggested by Eq. (4).

Accordingly, Eq. (4) becomes directly applicable to the characteristic H. and D. curves obtained and plotted in this manner. All exposures in the experiments to be described here were made in the manner outlined above, and the curves were drawn to the same abscissa scale. In applying Eq. (4) to these curves, a horizontal constant density line was extended across the characteristic H. and D. curves and the exposure values of the intersection points were read from the common abscissa scale for the red, green and blue curves, respectively. With these quantities, the exposure value for the mixture curve was calculated according to Eq. (4) and then compared directly with the observed value as read from the experimentally determined curve for the mixture of colors.

Results

The radiations used in these experiments, and referred to as R, G, and B, were obtained by use of a tungsten lamp and the Wratten tricolor filters, Red (No. 25), Green (No. 58) and Blue (No. 49). The yellow radiation used with the Azo paper was obtained with a Wratten 00 safelight. Development was carried out in D-90 developer for a period of four minutes at 68°F. All densities were read on a visual densitometer in which the comparison light beam was varied by means of a Martens polarization photometer.

The results of the experiments are presented in Table I. The calculated and observed values of $1/I_{r+g+b}$ are given in the column so labeled. Comparisons of the calculated and observed values of this quantity have been made at a series of density values for each emulsion tested.

TABLE I.

EMULSION	DEN.	$1/I_r'$	$1/I_b'$	$1/I_g'$	$1/I_{r+b+g}$ calc.	obs.	DIFF.	% ERROR
Super-	1.9	0.00276	0.00195	0.00229	0.00700	0.00725	−0.00025	−3.4
sensitive	1.6	.00513	.00398	.00427	.01338	.01335	+.00003	+0.2
Pan-	1.3	.00834	.00758	.00692	.02284	.02295	−.00011	−0.5
chromatic	1.0	.01320	.01445	.01048	.03813	.03890	−.00077	−1.9
Film	0.7	.02135	.02690	.01585	.06410	.06550	−.00140	−2.1
	0.4	.03720	.05380	.02750	.11850	.11770	+.00080	+0.7
	0.2	.07400	.12050	.04900	.24350	.23450	+.00900	+3.8
								−4.5 Av.
Super-	1.9	0.00276	—	0.00229	0.00505	0.00502	+0.00003	+0.6
sensitive	1.6	.00513	—	.00427	.00940	.00893	+0.00047	+5.3
Pan-	1.3	.00834	—	.00692	.01526	.01480	+.00046	+3.1
chromatic	1.0	.01320	—	.01048	.02368	.02345	+.00023	+1.0
Film	0.7	.02135	—	.01585	.03720	.03545	+.00175	+4.9
	0.4	.03720	—	.02750	.06470	.06170	+.00300	+4.9
	0.1	.07400	—	.04900	.12300	.1178	+.00052	+4.4
								+3.4 Av.
Commercial	2.2	—	0.00316	0.00398	0.00714	0.00678	+0.00036	+5.3
Ortho	1.8	—	0.00645	0.00692	0.01337	0.01350	−.00013	−0.91
Film	1.4	—	0.01318	0.01150	0.02468	0.02515	−.00047	−1.9
No. 1	1.0	—	.02600	.01970	.04570	.04470	+.00100	+2.2
	0.6	—	.05250	.03470	.08720	.08730	−.00010	−0.1
	0.3	—	.11500	.06920	.18420	.18200	+.00220	+1.2
								+1.0 Av.
Super	1.8	—	0.00339	0.01071	0.01410	0.01380	+0.00630	+2.2
Speed	1.4	—	.01000	.01995	.02995	.02890	+.00105	+3.6
Portrait	1.0	—	.02290	.03540	.05830	.05770	+.00060	+1.0
Film	0.6	—	.05250	.06760	.12010	.11770	+.00240	+2.0
	0.3	—	.12050	.13200	.25250	.25150	+.00100	+0.4
								+1.8 Av.
Panatomic	1.9	0.00398	0.00288	0.00288	0.00974	0.00955	+0.00019	+2.0
Film	1.4	.00742	.00645	.00537	.01924	.01950	−.00026	+1.3
	1.0	.01350	.01350	.00998	.03698	.03630	+.00068	+1.9
	0.6	.02750	.03390	.02090	.08230	.08320	−.00090	−1.0
	0.4	.05500	.07580	.04670	.17750	.17000	+.00750	+4.4
								+1.7 Av.
Portrait	1.8	0.00289	0.00364	0.02760	0.03413	0.03390	+0.00023	+0.7
Pan Film	1.4	.00575	.00914	.05380	.06869	.06770	+.00099	+1.5
	1.0	.01025	.02090	.09750	.12865	.12900	−.00035	−0.3
	0.6	.01820	.04780	.17800	.24400	.25150	−.00750	−3.0
	0.3	.03310	.10980	.34700	.48990	.52500	−.03510	−6.7
								−1.6 Av.
Verichrome	2.0	—	0.00662	0.00269	0.00932	0.00955	−0.00023	−2.4
Film	1.6	—	.01203	.00417	.01620	.01660	−.00040	−2.4
	1.2	—	.02090	.00662	.02752	.02820	−.00068	−2.4
	0.8	—	.03800	.01070	.04870	.04900	−.00030	−0.6
	0.4	—	.07240	.01995	.09235	.09550	−.00315	−3.3
								−2.4 Av.
Commercial	1.5	—	0.00282	0.01740	0.02022	0.01995	+0.00027	+1.3
Ortho	1.0	—	.00660	.03390	.04050	.03980	+.00070	+1.7
Film	0.5	—	.01660	.07250	.08910	.08520	+.00390	+4.6
No. II	0.2	—	.06320	.20000	.26320	.26300	+.00020	+0.1
								+1.9 Av.
Azo	0.95	0.0316*	0.0630	—	0.0946	0.1000	−0.0054	−5.4
No. I	.70	.0742	.0914	—	.1656	.1660	−.0004	−0.2
Paper	.50	.1202	.1202	—	.2404	.2400	+.0004	+0.1
	.30	.1900	.1695	—	.3595	.3630	−.0035	−0.9
	.10	.3240	.2690	—	.5930	.6025	−.0095	−1.6
								−1.6 Av.
Par	1.4	—	0.00467	0.00364	0.00831	0.00813	+0.00018	+2.2
Speed	1.0	—	.01070	.00795	.01865	.01820	+.00045	+2.4
Portrait	0.6	—	.02460	.01740	.04200	.03980	+.00220	+5.5
Film	0.3	—	.05250	.03470	.08720	.08330	+.00390	+4.7
								+3.7 Av.

* Yellow light.

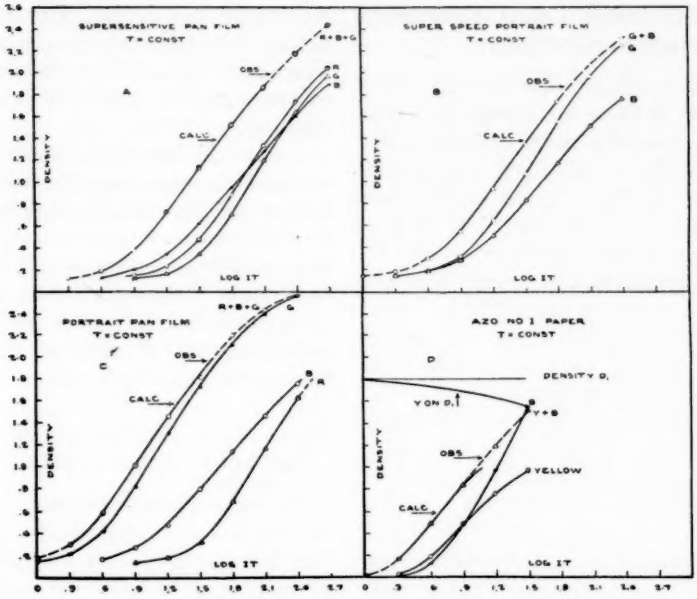


FIG. 1. Separate color curves; and mixture color curves (observed and calculated) obtained for the mixture of the different color radiations applied simultaneously.

The percentage differences between calculated and observed values are given in the last column, and they will give some idea of how accurately the theoretical Eq. (4) can predict the results of experiment.

The results for four of the different tests listed in Table I are shown graphically in Fig. 1. The experimental data are represented in all cases by the points shown. In the case of the H. and D. curve obtained with the mixture of colors, the *solid* line represents the theoretical curve calculated by means of Eq. (4) by using the data from the separate color curves. The agreement between this curve and the experimental points affords the test of Eq. (4). This graphical representation of the data has been included to emphasize the exactness with which the theoretical and observed values agree.

The curves of the (D) section of Fig. 1 are particularly to be noted. These exposures were made on Azo No. 1 paper, this material being chosen because of its ability to exhibit a very large Herschel effect. The radiations used in this case were a blue and a very intense yellow, the latter being capable of producing a large Herschel effect on the Azo paper. This is shown by the two curves labeled (D_1) and (Y on D_1). D_1 signifies a uniform density of 1.8 and the curved line below this indicates the reduction in density (Herschel effect) produced by an exposure to the same yellow radiation as used in obtaining the yellow H. and D. curve. Even in this case, the Van Kreveld law is seen to hold quite accurately up to the maximum value of density obtained with the yellow radiation. The summation law is, of course, not applicable beyond this point. It appears quite likely that, in this case, if the mixture exposure were allowed to continue indefinitely, the maximum density value might approach an equilibrium value predicted by the theoretical law.

A further very interesting aspect of the results of the above experiments is that, insofar as they confirm the validity of the Van Kreveld summation law, they show that the reciprocity failure curve of an emulsion is the same for all qualities of radiation, when compared on an equal time basis. A rigorous proof of this has been given by Van Kreveld[3] and so will not be repeated here.

[3] A. Van Kreveld, Physica 1, 60 (1934).

Addition of Different Color Radiations Applied Successively

In this part of the paper, experimental results will be presented to show the manner in which a photographic emulsion sums different color radiations when these radiations are applied *successively*. The results obtained have been employed to test the summation law of Van Kreveld in this case also. However, as will be seen later, the experimental procedure here was of necessity such that this law could be tested only in a rather degenerate form, the experimental restriction being imposed here by considerations of a fundamental nature.

The chief complication which arises in measurements of this type, is due to the reciprocity law failure. Unless the comparison exposures are carried out in such a manner that the resulting densities are produced at the same rates, differences due to reciprocity law failure will be introduced which will have to be taken into account in theoretical considerations. Eliminations of errors arising from this source in the case of the simultaneous addition of colors was obtained by making all the exposure curves *intensity scale* curves, for which the time factor was constant. The H. and D. curves, thus obtained, were compared point by point, by selecting those points for which densities were the same. Obviously points compared in this way were those for which the rates of reception of energy throughout the exposures were of such magnitude as to produce equal densities in equal over-all times for both the individual colors and the mixture of colors.

In the case of adding colors successively, however, this same procedure cannot be followed as the requirements are essentially different. The difficulty arises in connection with applying the two radiations of the mixture at intensity levels so that the *individual* radiations will produce densities at *equal rates*, and further, so that the total density produced by their combined action when applied successively will be equal to that produced by the separate radiations in the same over-all times. It is only on this basis that the action of the mixture of radiations and the action of the separate radiations can be compared with freedom from errors due to reciprocity law failure. This may be seen by noting that these are precisely the conditions which must be met if two identical radiations are to show perfect additivity when applied successively.

A method whereby these conditions have been met is here presented. The method consists of making the exposures by crossing intensity scale H. and D. curves to separate color radiations and of a graphical construction for comparison of densities of the separate color and the mixture color exposures. An explanation of the method and its theoretical significance will be outlined before presenting the results.

Theoretical Description of Experimental Method of Adding Colors Successively

The method used in making the experimental exposures is illustrated in the diagram of Fig. 2. The test piece of film was placed behind an intensity scale of radiation, with a single component color as source. In this way a series of exposure steps to one of the component radiations was impressed in one direction completely across the film. At the end of a specified time of x minutes, a dark slide containing a slot in the position $A-B$, was quickly thrust into position in front of the film and the exposure on the section, $A-B$, allowed to continue for a period

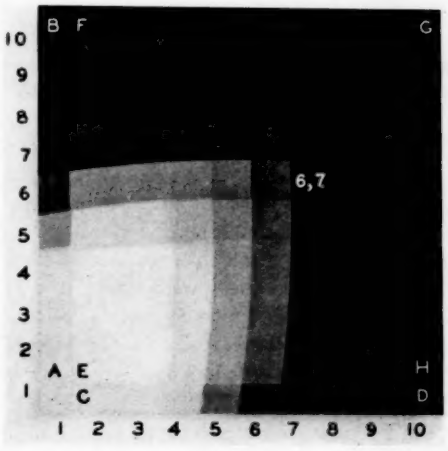


Fig. 2. Experimental arrangement of exposures for investigating the manner of addition of colors applied successively.

such that the total exposure time for the section $A-B$, was T minutes. Upon completion of this exposure, the test film was then rotated through 90° behind the intensity scale, and an exposure similar to the above made to the second component color radiation. The exposure steps in this case crossed the former exposure steps at right angles, as shown in the diagram. This second exposure was allowed to continue for a period of $(T-x)$ minutes, at the end of which time a dark slide with slot at $C-D$ was thrust into position in front of the film and the exposure on the portion of the film, $C-D$, allowed to continue for a period such that the total exposure time for this strip was also T minutes.

TABLE II.

FILM AREA	TYPE EXPOSURE	RADIATION	EXPOSURE TIME
$A-B$	Step intensity scale	Component 1	T min.
$C-D$	Step intensity scale	Component 2	T min.
$=EFGH$	Step intensity scale	Component 1	x min.
$\|\|EFGH$	Step intensity scale	Component 2	$(T-x)$ min.

Thus, it may be seen that the strip, $A-B$, received an intensity scale exposure of T minutes to one component color radiation, strip $C-D$, an intensity scale exposure of T minutes to the second component color radiation, and the area $EFGH$, a mixed exposure of total time T minutes, composed of x minutes to the first component color radiation and $(T-x)$ minutes to the second component color radiation. These exposures have been tabulated above for convenience. In all of the experiments performed, T had the value 16 minutes and x the value 4 or 8 minutes.

The purpose of this arrangement of exposures was to make it possible to select from the area, $EFGH$, which received the exposure to the mixture of radiations, that small area (e.g., square 6, 7), which was exposed to those intensities required to produce equal densities in equal times for the individual component radiations. Thus, suppose the square 6 of strip, $A-B$, should be found to have a density D_1, and square 7 of strip, $C-D$, likewise a density, D_1. Then the square 6, 7 of the area, $EFGH$, in this case would have been exposed for a total period of T minutes to the component radiation intensities [$T-x$ to component 1, and x to component 2] either of which alone produced a density, D_1 in T minutes. Thus, the *effective* intensity throughout the mixed exposure might be considered constant and equal for the two component colors. If, under these conditions, the square 6, 7 should also be found to have a density D_1, then there would be perfect additivity of the colors by the emulsion. This criterion of perfect additivity is thought to be a logical one, since this would be the result if the component color radiations were identical type radiations.

In the actual experiments, the exposure curves were obtained by varying the intensity values in finite increments, as illustrated in Fig. 2. It was therefore necessary, in general, to interpolate between steps on the exposure curves in order to carry out the comparison of densities as above outlined. It was found that this comparison could be accomplished most readily in practice by a graphical process. For this purpose the curves, $A-B$, and $C-D$, were plotted directly as illustrated in Fig. 3. In order to obtain a continuous graphical representation of the experimental density values in the area, $EFGH$, the measured density values of the steps in this area were written in numerical form at the corners of a square mesh system, as shown in the diagram. With these values, the iso-density lines a, b, and c, were derived and plotted as illustrated. Only three of the iso-density lines, corresponding to the density values D_1, D_2, and

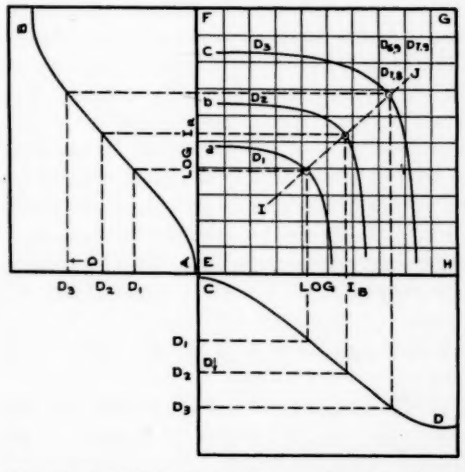


FIG. 3. Graphical construction employed to show the manner in which a photographic emulsion adds different color radiations applied successively.

D_3 are shown here. The theoretical points, which are to be compared with the experimental iso-density curves, were obtained from the H. and D. curves for the separate component radiations in the following manner. Choosing equal density levels, as D_1, on the two H. and D. curves, lines corresponding to the log I values for these densities were projected, as shown by the dotted lines, into the area of the iso-density curves. The intersections of these lines locate the theoretical points, designated in the figure by dots surrounded by circles. If the emulsion adds the two component radiations perfectly, i.e., in the same way that it would add two radiations of the same quality, then these theoretical points should fall on the experimentally determined iso-density curves *a*, *b*, and *c*.

In presenting the experimental results of this investigation, the general scheme outlined above has been followed. In order to make the graphs as compact as possible, Fig. 3 has been deviated from to the extent of placing the H. and D. curves within the square, *EFGH*, and of confining the iso-density curves to a region between these curves. The iso-density curves have in all cases, however, been extended sufficiently to include the theoretical points derived from the H. and D. curves.

Apparatus and Experimental Procedure

In making the experimental exposures in this work, as also in that on addition of colors simultaneously, the intensity scale curves were obtained by means of a rapidly rotating logarithmic sector. The exposures being of considerable duration (16 minutes), a moderate rate of rotation was found to be sufficient to yield curves in agreement with those obtainable with step wedge or other standard methods of making intensity scale curves. The logarithmic sector used was that belonging to the reciprocity sensitometer, described elsewhere,[4] and the speed of rotation employed was 16 interruptions per second.

The light source consisted of a lamphouse with ground-glass window. Inside the lamphouse was placed a 250-watt tungsten lamp maintained at a color temperature of 2500°K. In front of the ground-glass window was placed a metal diaphragm containing three holes of approximately one-inch diameter arranged in triangular form as shown in Fig. 4. Over the three windows were placed red, green, and blue filters, respectively. The filters used were the tricolor Wratten filters: Red, No. 25, Green, No. 58, and Blue, No. 49. Neutral densities were used in conjunction with these filters to regulate the intensities of each of these sources to an appropriate value. Appropriate value here means intensity values, which were nearly equal in their photographic effect for the emulsion under test.

The exposures were carried out in accordance with the method outlined in the preceding section. In making the exposures to the different colors, the elapsed time between was reduced to the minimum compatible with the rearrangement of the film, in order to minimize any growth effects of the latent image. The arrangement of the exposures on the film is shown in Fig. 2. It may be seen that it was not possible with the method of experiment employed here to study the effect of superposition of more than two component radiations on a single emulsion. This becomes obvious by noting that it is impossible to cross exposures from three or more sources in the manner outlined, so that there will be in the mixed exposure area a *single* point corresponding to the intensity values for which equal densities are produced by the individual radiation of different colors. The procedure adopted in this

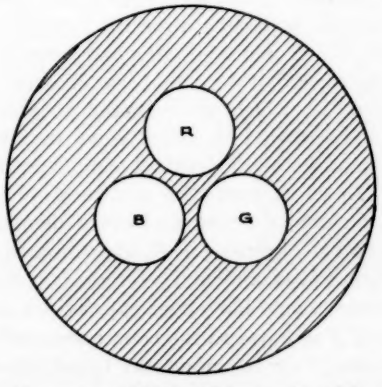


Fig. 4. Arrangement of the different color light sources.

[4] L. A. Jones, J. Opt. Soc. Am. & Rev. Sci. Inst. **1**, 305 (1923).

work for testing an emulsion for more than two colors consisted in working with two colors at a time. If it were desired to test an emulsion for three colors, e.g., as *a*, *b*, and *c*, exposures were made to the pairs of colors (*a*,*b*), (*b*,*c*) and (*a*,*c*) separately. If the emulsion summed each pair of colors correctly, according to the criterion already stated, then it was considered legitimate to assume that the emulsion would also sum the three radiations correctly.

The distance and source intensity of each of the windows of the lamphouse were kept the same during each experiment. The total times of the exposures were the same in all experiments (16 minutes). The exposure times for the individual color radiations in the mixed exposure area were varied somewhat in different experiments, but always so that the combined exposure times to the two colors totaled 16 minutes. Component 1 was given x minutes (either 4 or 8 minutes) and component 2 was given $(16-x)$ minutes.

Since each experiment was placed on a piece of film four inches square, all the exposures of a single experiment could be developed simultaneously. Development was carried out in the Eastman D-90 developer for a period of four minutes, at 68°F.

Densities were read on a visual densitometer, which employs a Martens polarization photometer for varying the intensity of the comparison beam of light.

Results

The results of the experiments are shown graphically in Figs. 5 to 7. The information necessary for the interpretation of the results has been included on these diagrams. The two H. and D. curves, density *vs.* log I (I in arbitrary units), are portrayed graphically in each figure. These curves are shown as a function of the intensity factor of exposure in order to emphasize the fact that the curves are intensity scale curves, and further, because if plotted in this way, the same intensity values also apply to the densities in the mixed exposure area. The time factors of the exposures are given in the legend at the upper left corner of the figures. For example, in Fig. 5, the notation, $T_r=16$M & 8M, signifies that the exposure time for the red H. and D. curve was 16 minutes, and the exposure time for the red part of the exposure in the mixed exposure was 8 minutes. The observed densities in the mixed exposure area are given by the numerical values placed at the corners of the squares of the mesh system. From these values the iso-density curves, shown as solid lines, were derived. The black dots signify the interpolated points used in constructing these curves. These iso-density curves indicate the relationship between the observed values of intensities of the component radiations required to produce a given constant density in the mixed exposure area. The points indicated by dots with circles, specify the theoretical values of the intensities of the component radiations required to produce this same density on the basis that the emulsion adds the two radiations correctly. Therefore, the closeness with which these points fall on the iso-density curves may be taken as a measure of how nearly correctly the emulsion adds the two component radiations.

The summation of colors for each emulsion was tested for a number of density values. Iso-density curves for values 0.3, 0.5, 1.0, 1.5, and 2.0 were constructed in all cases, and additional curves for values above and below this range were included where possible.

The type of film and the color of the radiations employed in each experiment are given on the diagram. The colors of the different radiations are indicated by a capital letter, and their order, e.g., as *G* on *R*, signifies the order in which the different color exposures were made.

An inspection of Figs. 5 to 7 will show that, in all cases, the emulsions added the different color radiations with a high degree of accuracy. In only a few cases do the theoretical points deviate appreciably from the iso-density curves, and in all such cases, the deviations lie within the limits of experimental error in this work.

Discussion of Results

Of chief theoretical significance in the present experimental work is the result that an emulsion sums radiations of different quality in the same manner as it sums radiations of identical quality. This result is considered as supporting evidence

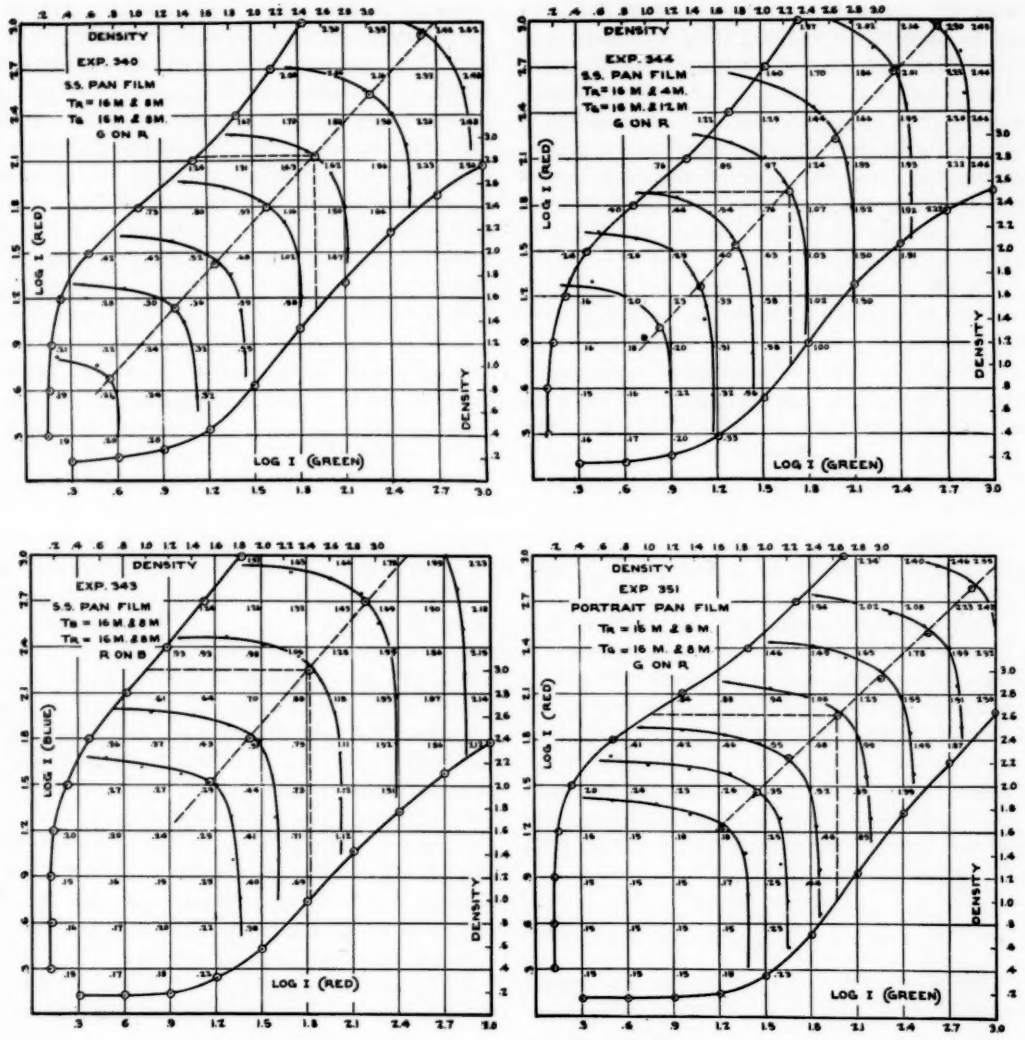


FIG. 5.

of the view that the same photographic grain must be affected indiscriminately by different color radiations. This may be seen from the following consideration: Since the grain is the exposure receiving unit of the emulsion, the reciprocity law failure (variation of density with intensity at constant exposure) must be governed by the rate of utilization of energy by the individual grains. Since, as shown by these experiments, the same conditions must be met in order to eliminate reciprocity law failure in the case of mixtures of different color radiations as in that of mixtures of identical radiations, it is thought safe to conclude that individual grains are affected by the different color radiations. The results of these experiments show further that an emulsion cannot be sensitized for one radiation by pre-exposure to another radiation.

As mentioned earlier, the summation law was tested in the present experiments in which the

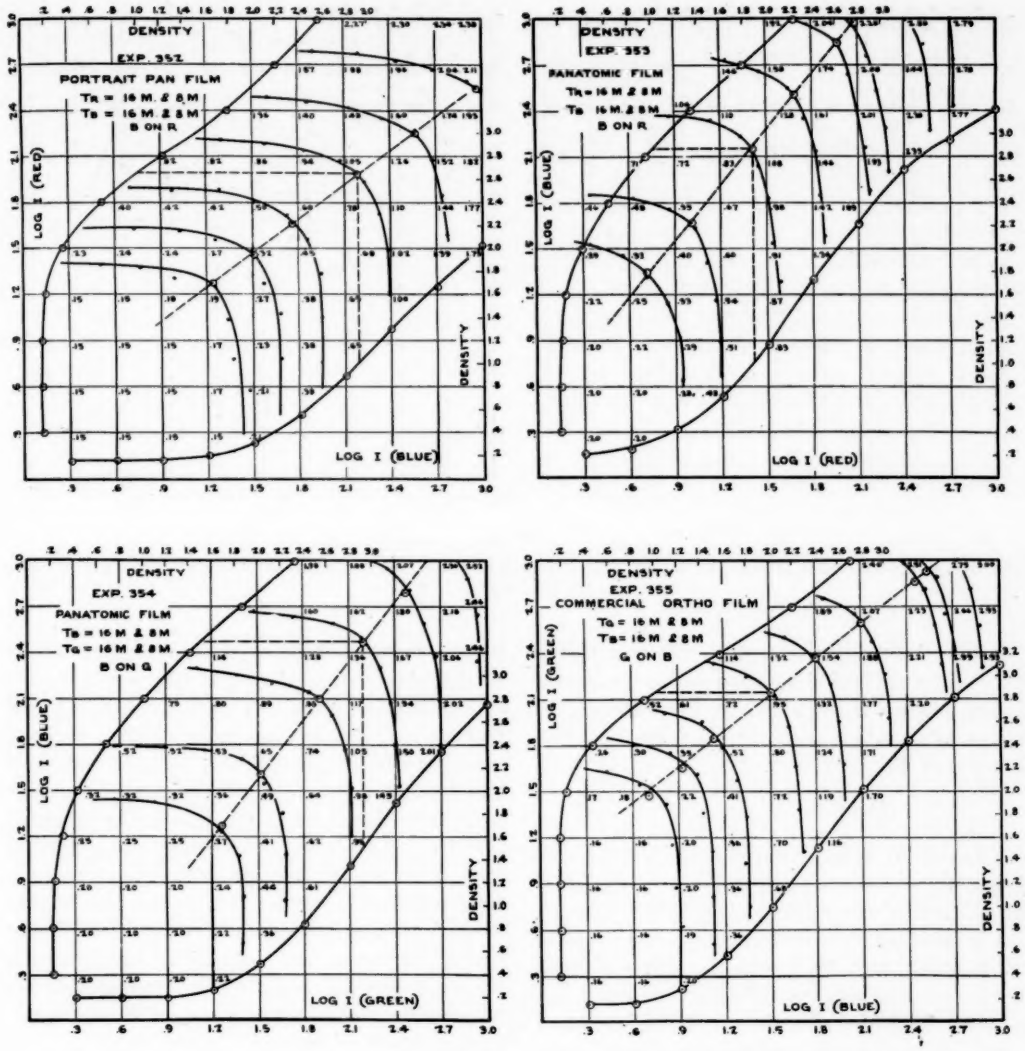


FIG. 6.

colors were applied successively. It will be shown that under the restricted conditions of these experiments, this law also holds for this case. For the case of adding colors successively, it is more convenient to write the summation law (1) in the slightly altered form:

$$\left[\frac{I_m{}^r t_m{}^r}{I_r t_r}+\frac{I_m{}^g t_m{}^g}{I_g t_g}+\frac{I_m{}^b t_m{}^b}{I_b t_b}=1\right]_{\text{den.=const.}}, \tag{5}$$

in terms of the intensities and times explicitly. In this equation, $I_m{}^r$ and $t_m{}^r$ signify the intensity and time factors, respectively, of the red part of the mixed exposure. As the exposures were carried out under the special conditions, $I_m{}^r = I_r$, $I_m{}^g = I_g$, $I_m{}^b = I_b$, etc., to eliminate errors of reciprocity failure, the Eq. (5) reduces to the form:

$$[t_m{}^r + t_m{}^g + t_m{}^b = t_r = t_g = t_b]_{\text{den.=const.}}. \tag{5'}$$

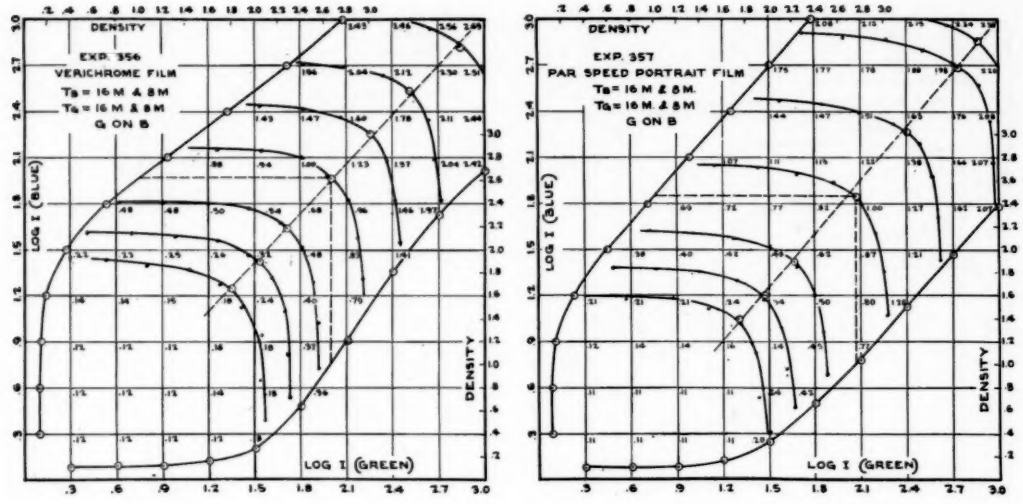


FIG. 7.

FIGS. 5–7. Graphs showing the manner in which a photographic emulsion adds different color radiations applied successively.

It may be readily seen that this equation is satisfied for each of the comparison experiments shown in Figs. 5 to 7. Taking for example the values of t_r, $t_m{}^r$, t_g, and $t_m{}^g$ from the S.S. Pan diagram of Fig. 5, the result

$$(4+12=16)_{\text{den.}=\text{const.}}$$

is obtained. It is to be noted here that the condition, density = constant, furnishes the experimental test of the addition law. The procedure is a reversal of that used in the case of adding colors simultaneously. There, the points of equal density were chosen and comparison was made between the predicted and experimental exposure values. Here, however, the exposures were made in accordance with the theoretical equation and the test consisted of the comparison of the densities for the mixture and separate color exposures.

It is seen that the addition law has been confirmed under very restricted conditions when colors are applied successively. However, as the restrictions in this case are imposed by the reciprocity law failure it is thought that this restricted form of the law (Eq. (5′)) is the most general form that can be said to apply for such measurements.

An interesting feature of the foregoing results is the observation that the theoretical points established in Figs. 5 to 7 were in all cases found to lie on a straight line. At first sight this appeared to be accidental, but the persistence of this feature throughout all the diagrams seems to indicate that it is not. Obviously, this straight-line feature of the points must exist in that region for which both H. and D. curves are straight. But the extension of this result into the over- and underexposure regions of the curves indicates a relationship between the shapes of the curves that is not apparent from casual observation. On the basis of this observation, a simple relationship, between the slopes of the H. and D. curves for different colors can be derived. Looking back to Fig. 3, the theoretical points are shown along the straight line, IJ, in accordance with the experimental results. From inspection, the following relationship is immediately seen to hold between the log I_r and log I_b values for points along this line;

$$(\log I_r = \log C + M \log I_b)_{\text{points of equal density.}} \quad (6)$$

In the expression (6), r and b refer to the colors of the two component radiations used, (log C) is a constant, which may be made equal to zero

by a shift of coordinates, and M is the slope of the straight line, IJ. By means of Eq. (6), it is possible to predict the relationship between the slopes (contrast values) of the (R) and (B) H. and D. curves at each level of density. To demonstrate this, it is necessary to write Eq. (6) for two neighboring values of density, as:

$$\begin{aligned}(\log I_r^1 &= \log C + M \log I_b^1)_{\text{density}=D_1},\\ (\log I_r^2 &= \log C + M \log I_b^2)_{\text{density}=D_1+\Delta D}.\end{aligned} \tag{7}$$

Subtracting the second equation from the first, the result is:

$$\Delta(\log I_r) = M\Delta(\log I_b). \tag{8}$$

By taking the reciprocal of Eq. (8) and multiplying both sides by ΔD, the expression

$$\left[\frac{\Delta D}{\Delta(\log I_r)} = \frac{1}{M}\frac{\Delta D}{\Delta(\log I_b)}\right]_{\text{points of equal density}} \tag{9}$$

is obtained for the relationships between the slopes of the two H. and D. curves at each density value.

The above relationship probably results from the fact that the slope of an H. and D. curve is largely determined by the grain size distribution in the emulsion under experiment. On this basis, when exposures are made on the same emulsion to two different colors, this common feature (grain size distribution) manifests itself in this way.

Refraction of Tempered Plate Glass

JOHN J. HOPFIELD AND GEORGE B. WATKINS, *Libbey-Owens-Ford Glass Company, Toledo, Ohio*

(Received October 16, 1935)

It was found that the refractive index of heat-strengthened plate glass diminishes continuously from surface to center. As a consequence (1) prisms of such glass form spectral lines that are greatly curved, (2) plates of glass with plane parallel ends refract like cylindrical, concave lenses. This change of refraction is probably due to the condition of strain in the glass rather than to a relative change in the concentration of the component oxides.

THE double refraction of light by strained glass is well known, and in well annealed glass, it is almost entirely absent.

Heat-strengthened glass contains such a distribution and amount of strain that some of its properties are found to be greatly modified. For example, quarter-inch plate glass of ordinary window size will withstand much greater loads and resist much greater thermal shock than annealed glass.

The heat-strengthening operation consists in placing pieces of the glass, cut to the desired size and shape, into a hot furnace where the glass is heated rapidly but uniformly just to the point of softening. They are then withdrawn from the furnace and cooled rapidly and uniformly by streams of cold air. It is obvious that this rapid cooling freezes first the outside surface of the glass while the inside of the glass is still dilated and plastic. When subsequently the inside freezes, it tends to shrink away from the surface shell already formed. As a result, the surface layers of the glass are under compressional strain and the central portion is under tension.

In measuring the refractive index of prisms of this glass in comparison with prisms of the same glass when untreated and in comparing blocks of both glasses, striking differences were found. The results are shown in part in Figs. 1 and 2.

Fig. 1 is the spectrum of a long, glass mercury arc taken through two thin (one-fourth inches high) 60° prisms of glass. The top prism was of annealed glass and the lower one of heat-strengthened glass. The figure shows that the spectrum of the top prism is usual in that it consists of straight spectral lines, while the heat-strengthened glass prism gives greatly curved lines. The comparison of these two forms of spectra indicates (1) that the refractive index of the surface layers of the heat-strengthened glass is nearly equal to (a little less than) the refractive index of the annealed glass, and (2) the refractive index of the central layers is considerably *less than at the surface.* This diminution of the refractive index towards the center of the glass indicates that the change is not primarily one due to an evaporation of the more volatile constituents of the glass leaving a higher concentration of silica, for if this were the chief cause, then the change in the refraction would be the reverse of that observed.

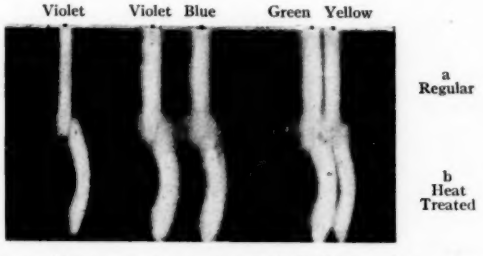


FIG. 1. Spectrum of mercury. (a) Ordinary glass prism; (b) heat-strengthened prism.

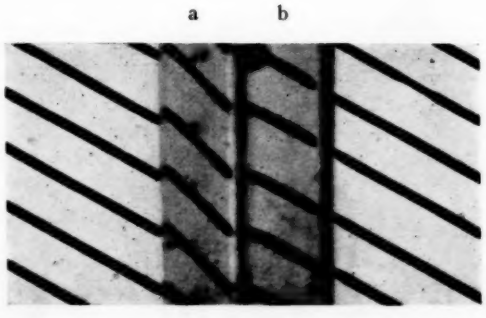


FIG. 2. Cylindrical lens effect. (a) Heat-strengthened glass; (b) slab of ordinary glass.

Some interesting things as to the optical action of heat-strengthened glass may be predicted from its spectral behavior as outlined above. For example, a plate of the above glass with plane parallel ends should behave like a cylindrical concave lens for objects viewed through these ends. Fig. 2 shows the action of a piece of heat-strengthened glass (a) and a piece of annealed glass (b) held side by side and above some oblique lines on a piece of paper. The distance between the viewing ends was about two inches. It is seen that glass (b) behaves normally, but that glass (a) tilts the oblique lines seen through it just as would a cylindrical concave lens whose cylindrical axis is vertical. (The reader can easily check this cylindrical refraction on his own spectacles if they have astigmatic correction.) A further tilting of the lines with reference to *A* results in a doubling of the ends of the lines. These are obviously the ordinary and the extra-ordinary ray components of the transmitted light.

Further predictions of these data that have not yet been tested are: That a "Prince Rupert Drop" should have a greater focal length than a drop of the same size made of untempered glass. That an approximately spherical concave lens could be made from a normal plane parallel segment of a round cylindrical rod of tempered glass. That in "annealed" plate and sheet glass a slight stratification of refractive index should be observable.

More important than any of these is the possibility of using this stratification of refraction in studying the large strains in tempered flat glass.

Pressure Effect on Sensitivity of Different Photographic Films

NY TSI-ZÉ AND LÜ TA-YUAN, *Institute of Physics, National Academy of Peiping*

(Received August 12, 1935)

The effect of pressure on photographic sensitivity is studied with the monochromatic radiations of a quartz mercury arc for four different films of Eastman Kodak Company; namely, Kodalith film, motion picture positive film, commercial film, and portrait film. The quantity E/E_e, where E is the illumination actually used when the emulsion is under pressure, and E_e the illumination which would be required to produce the same optical density when the emulsion is under no pressure, is a function of pressure only and, for pressures greater than 250 kg/cm², E/E_e is a linear function of the pressure. The variation of E/E_e with the wave-length of the radiations is very similar to that of the contrast for all the films studied. It is pointed out that the pressure effect on photographic sensitivity may be closely related to the absorption properties of the emulsion.

INTRODUCTION

THE application of pressure to a photographic emulsion during exposure prevents the formation of latent image to a great extent and thus reduces its sensitivity.[1] To characterize this effect, one of us[2] has introduced a quantity E/E_e, E_e being the illumination which would produce, on an emulsion without being subject to any pressure, the same optical density as that produced by illumination E on the same emulsion under a pressure p with equal time of exposure. E_e is always less than E and may be called in photography the effective illumination of E under pressure p. The quantity E/E_e is a function of pressure p only, and for p greater than a certain value, the ratio E/E_e is a linear function of the pressure.

It has also been shown[3] that the effect of pressure on photographic sensitivity varies very much with the wave-length of the radiation used. It diminishes, in general, with the wave-length. For the yellow or green light a pressure of less than one hundred kilograms per square centimeter is sufficient to produce a marked effect, while for the radiation 3131A, we have to apply more than one thousand kilograms per square centimeter in order to make the effect just observable.

The pressure effect on photographic sensitivity differs from one kind of film to another.[4] It is proposed in the present paper to investigate systematically this effect upon materials differing widely in characteristics, and to correlate it, if possible, with the known properties of the different films, so as to throw some light on the nature of this effect.

EXPERIMENTAL ARRANGEMENT AND METHOD

The experimental arrangement is schematically shown in Fig. 1. The light of a quartz mercury lamp is projected by a quartz condenser C on the slit S_1 of a Leiss monochromator.[5] After entering the slit S_1, the light is rendered parallel by the spherical mirror M_1, passes through a Cornu prism P at minimum deviation, is reflected by a Wadsworth plane mirror M, and

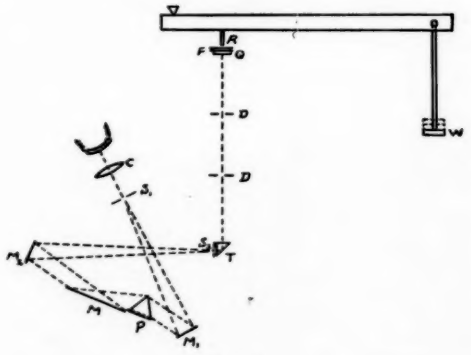


FIG. 1.

[1] P. E. Poindexter, J. Opt. Soc. Am. **2**, 59 (1931).

[2] Ny Tsi-Zé and Tsien Ling-Chao, Comptes rendus **194**, 1644 (1932); Science et industries photographiques (2), **4**, 1 (1933).

[3] Ny Tsi-Zé and Tsien Ling-Chao, Comptes rendus **196**, 107 (1933).

[4] Ny Tsi-Zé and Tsien Ling-Chao, Chinese J. Phys. **1**, 66 (1934).

[5] C. Leiss, Zeits. f. Physik **72**, 822 (1931).

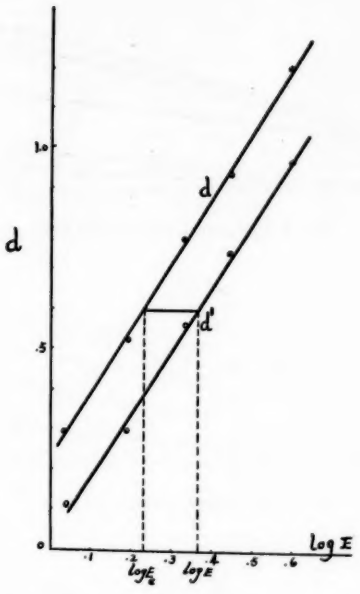


FIG. 2. d log E and d' log E curves of positive film under a pressure of 716 kg/cm² for 5791–70A, $E/E_e = 1.38$.

finally converged by a second spherical mirror M_2 on the slit S_2. On rotating, as a whole, the prism and the mirror M which remains always parallel to the base of the prism, any line of the mercury lamp can be brought before the slit S_2. The beam of monochromatic light through S_2 is then totally reflected upwards by a quartz prism T.

The photographic film was placed on top of a quartz plate Q with its sensitive layer next to the plate. The pressure was applied by means of a lever and weights W which forced the well flattened base of a steel cylinder R against the back of the film F. The light from the slit S_2 which was about 65 cm from the film was confined, by means of the two diaphragms D, to a beam circular in cross section having a diameter of 14 mm on the quartz plate Q, while the flattened base of the cylinder had a diameter of 5 mm and was approximately concentric with the beam of light. To avoid reflection and other possible effects due to direct contact, a sheet of black paper or an exposed and developed film was interposed between the base of the metallic cylinder R and the photographic film under investigation.

The experiment was performed on one film, either with the same pressure but different radiation, or with the same radiation but different pressures. For a certain radiation and one given pressure, five normal exposures of equal time with different intensities were always made, the intensity of illumination being varied by means of screens inserted in the path of light. These screens consisted of exposed and developed photographic sensitive layer between two quartz plates, and their densities had been previously calibrated.

The measurement of the densities at the center and the rim of the photographic images with a Moll microphotometer permits one to trace the two curves $d \log E$ and $d' \log E$ in Fig. 2, which are parallel to each other. The translation parallel to the axis of abscissa bringing about the coincidence of the curve d with the curve d' measures the quantity $\log E/E_e$ and consequently the ratio E/E_e.

The time of exposure varied from ten seconds to three minutes according to the intensity of the different mercury lines and to the sensitivity of the different films.

RESULTS

Four different films of Eastman Kodak Company[6] were used in this experiment. They are: Kodalith film, motion picture positive film, commercial film, and portrait film. They represent a very wide range in speed and grain size frequency characteristics. All these films have their maximum sensitivity at about 4600A, and the portrait film has another secondary maximum at 5600A.

The pyro developer (formula D-90) recommended by Eastman Kodak Company was used. The temperature of the developer and the time of development were kept constant throughout our experiment so that the contrast γ of each film for different radiations of the mercury arc can be measured and compared.

The results obtained with these four films for

[6] These films were supplied to us through the kindness of Dr. Loyd A. Jones, to whom we express here our heartiest thanks and appreciation.

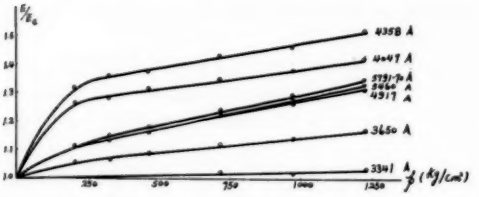


FIG. 3. Eastman regular Kodalith film.

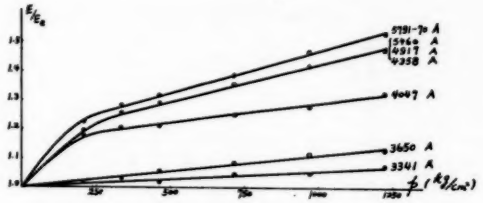


FIG. 4. Eastman Ciné positive film.

TABLE I.

Pressure p (kg/cm²)	5791 / 5770	5460	4917	E/E_e 4358	4047	3650	3341	3131A
EASTMAN REGULAR KODALITH FILM H AND D SPEED: $10/i=0.525$.								
208	1.11	1.11	1.11	1.32	1.27	1.06		No effect
335	1.15	1.14	1.14	1.36	1.29	1.07		
462	1.18	1.17	1.17	1.38	1.32	1.09		
716	1.24	1.22	1.22	1.43	1.35	1.12	1.02	
970	1.29	1.28	1.27	1.47	1.38	1.14	1.02	
1224	1.35	1.33	1.32	1.52	1.42	1.17	1.03	
EASTMAN CINÉ POSITIVE FILM H AND D SPEED: $10/i=50$.								
208	1.23	1.19	1.19	1.19	1.17	1.02		No effect
335	1.28	1.25	1.26	1.25	1.20	1.03		
462	1.31	1.28	1.28	1.29	1.21	1.05	1.02	
716	1.38	1.35	1.34	1.34	1.25	1.08	1.04	
970	1.46	1.41	1.41	1.41	1.28	1.11	1.05	
1224	1.53	1.48	1.48	1.47	1.32	1.13	1.07	
EASTMAN COMMERCIAL FILM H AND D SPEED: $10/i=269$.								
208	1.25	1.23	1.21	1.13	1.06	1.02		No effect
462	1.33	1.31	1.30	1.22	1.14	1.05		
716	1.40	1.37	1.35	1.28	1.16	1.07		
970	1.45	1.44	1.39	1.32	1.20	1.09		
1224	1.50	1.48	1.44	1.36	1.27	1.11		
EASTMAN SUPER-SPEED PORTRAIT FILM H AND D SPEED: $10/i=646$.								
208	1.41	1.41	1.21	1.17	1.09	1.04		
335	1.54	1.54	1.30	1.24	1.12	1.06		
462	1.64	1.63	1.38	1.31	1.16	1.08	1.03	
716	1.76	1.76	1.50	1.40	1.21	1.11	1.04	
970	1.90	1.90	1.58	1.47	1.27	1.16	1.06	
1224	2.02	2.03	1.69	1.57	1.32	1.21	1.08	1.03

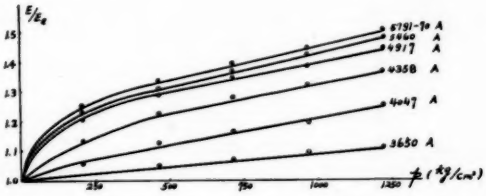


FIG. 5. Eastman commercial film.

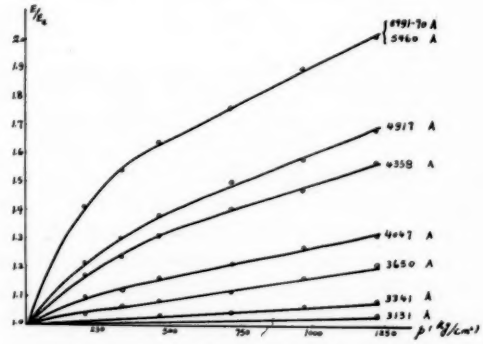


FIG. 6. Eastman super-speed portrait film.

different radiations under different pressures are given in Table I. With p as abscissa and E/E_e as ordinates, we obtain for each film a family of curves (Figs. 3–6), which are rectilinear for p greater than 250 kg/cm². The portrait film, which is the fastest among the four films studied, has a greater pressure effect than the others; but the positive film, of which the H. and D. speed is about five times smaller than that of the commercial film and a hundred times greater than that of the Kodalith film, has an effect nearly equal to these two. It is thereby inferred that the effect of pressure on the sensitivity is not simply related to the speed or the grain size frequency characteristics of the emulsion. This is in agreement with the results[7] relative to the Zeiss-Ikon orthochromatic film and the Agfa ultrarapid isochromatic film.

The effect of pressure on photographic sensitivity diminishes with the wave-length of the radiation used, except in the case of Kodalith film which presents a maximum effect at 4358A. This pecularity of the Kodalith film may be approached with the fact that this film presents

[7] Cf. Ny Tsi-Zé and Tsien Ling-Chao, Chinese J. Phys. 1, 66 (1934). These two films are both very rapid, but the Zeiss-Ikon orthochromatic film shows a pressure effect greater than that of the portrait film, while the Agfa isochrome film shows an effect even smaller than that of the Kodalith film.

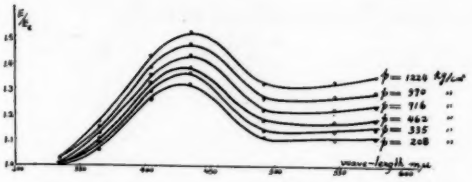


FIG. 7. Eastman regular Kodalith film.

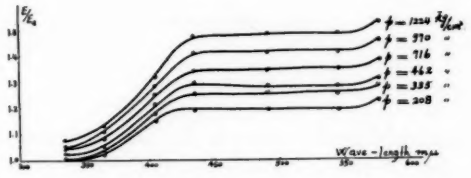


FIG. 8. Eastman Ciné positive film.

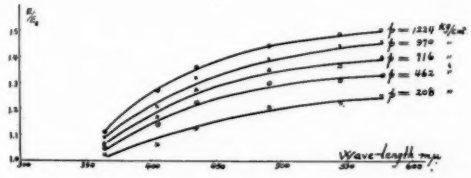


FIG. 9. Eastman commercial film.

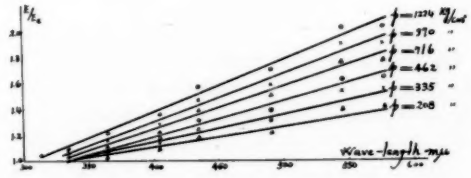


FIG. 10. Eastman super-speed portrait film.

also a maximum contrast in the neighborhood of 4358A.

To bring out this point more prominent we plot E/E_e against λ (Figs. 7–10) for each pressure with the data given in Table I. Values of gamma of different radiations for various films determined at the same time as E/E_e from the characteristic curves of Fig. 2 are shown in Fig. 11.

The E/E_e–λ and γ–λ curves for each film are very similar in form. Hence it seems to us that the variation of pressure effect on sensitivity with wave-length will be closely related to the

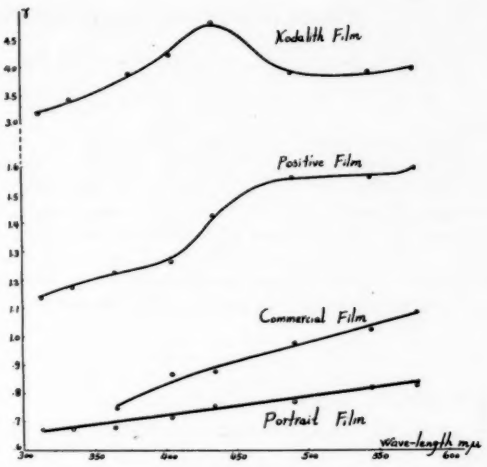


FIG. 11.

variation of gamma with the wave-length and consequently to the absorption properties of the photographic sensitive layer for different radiations.

In fact, Fabry and Buisson[8] have pointed out the great absorption of the sensitive layer for the ultraviolet and established for the contrast:

$$\gamma = m'/m,$$

where m is the absorption coefficient of the emulsion for the radiation considered and m' that of the completely reduced emulsion for the visible radiations used in the ordinary measurement of optical density. It follows that under the action of ultraviolet, the sensitive layer of a photographic plate of film is only superficially affected, while for the visible light, the photographic impression is produced throughout the whole thickness. By the similarity of the general form of the curves E/E_e–λ and γ–λ, we are therefore inclined to admit that the pressure would affect the internal layers of the emulsion more than the surface layer. It seems easy to account for this experimental result on a crystal-cracking hypothesis of latent image formation; that is, the pressure prevents the vibrational response of the molecules to the radiation, especially the molecules in the deep layers of the emulsion.

[8] Fabry and Buisson, Rev. d'optique **3**, 19 (1924).

Effect of Oxygen Upon the Photoelectric Thresholds of Metals

H. C. RENTSCHLER AND D. E. HENRY, *Research Laboratory, Westinghouse Lamp Co., Bloomfield, N. J.*

(Received July 13, 1935)

When a small amount of oxygen is allowed to react with an active surface of certain metals such as thorium, uranium, calcium, barium and caesium the photoelectric threshold is shifted toward the longer wave-length. A similar reaction of oxygen with other metals such as titanium, zirconium, silver, iron, nickel, etc., shifts the threshold toward shorter wave-length, while oxygen produces no effect on gold. The photoelectric work functions for the oxygen treated surfaces of thorium, uranium and calcium obtained agree with the thermionic work functions of tungsten filaments containing oxides of these metals when activated. The experimental results indicate that the effects from these oxygen treated surfaces are due to the activity of the product of the interaction of oxygen with the metal and not due to a single atom layer of the metal on a base material.

IN 1932 a simple method was described for preparing photoelectric cells of practically any metal.[1] Characteristic curves for different cells of the same metal were shown to be remarkably uniform, indicating that the method is capable of giving true photoelectric values of the metal in question. The photoelectric work functions for tungsten and tantalum as determined from the photoelectric response curves of these metals were found to agree with the most reliable values available for their thermionic work functions.

Recently this work has been extended to other metals and the following threshold wave-lengths determined:

Molybdenum	approx. 2850A
Platinum	less than 2000A
Iron	approx. 2680A
Nickel	2550A
Chromium	2840A
Bismuth	2870A

The corresponding photoelectric work functions calculated from the threshold wave-lengths for molybdenum and platinum are 4.33 and greater than 6.17 volts, respectively, agreeing again with the thermionic work function 4.31 volts for molybdenum, obtained by Dushman,[2] and the photoelectric and thermionic work functions of 6.30 and 6.27 volts for platinum, determined by DuBridge.[3] This agreement of the photoelectric work functions for tungsten, tantalum, molybdenum, and platinum with the thermionic work functions for the same metals further shows the reliability of the method for producing surfaces from which the true photoelectric constants may be determined.

In the original article[1] the photoelectric response for the material evaporated from an activated oxide coated cathode was shown to contain a constituent which has a considerably lower work function than that of pure barium. To determine whether such increased activity can be produced by the action of oxygen upon barium and other metals the following tests were made.

APPARATUS

Photo-cells of the different metals were prepared and tested as previously described.[1] Small amounts of oxygen were then introduced into the cell in the following manner. As shown in Fig. 1, a tube *T* was sealed to the cell before exhaust (Fig. 1). A small glass tube *C* drawn to a point at one end was filled with pure dry oxygen at a definite pressure and inserted in the side tube *t*. The cell was tested for its photoelectric response to different wave-lengths from a mercury arc as previously described.[1] An iron plunger *F* was then raised and allowed to drop and break the tip of the oxygen tube *C*. A reaction between the active surface and the oxygen was evident from the change in the cell response to the different wave-length radiations. To assure strictly comparable conditions, on final test for the effect produced by the oxygen, an iron plunger was inserted in the tube *T* above *B* and the cell was again sealed to a high vacuum pump at *P*. After the part above *B* was thoroughly exhausted, the iron plunger was

[1] Rentschler, Henry and Smith, Rev. Sci. Inst. **3**, 794 (1932).

[2] Dushman, Phys. Rev. **21**, 623 (1923).

[3] DuBridge, Phys. Rev. **29**, 451 (1927); **31**, 236 (1928).

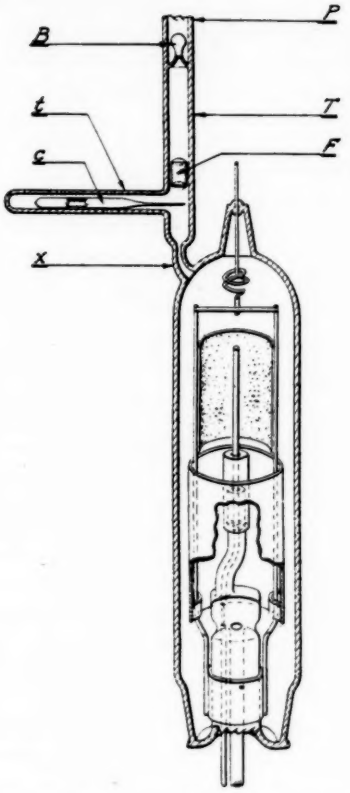


FIG. 1. Photo-cell with side tube for introducing a definite small amount of oxygen.

raised by a magnet and allowed to drop and break the bulb *B*. Uncombined oxygen, if any, that remained in the cell was again pumped out, the cell sealed off at *X* and again tested. To determine the effect of more oxygen several side tubes as *t* with oxygen containers *C* were attached to the main tube *T* using additional bulbs as *B* for successive re-exhausts.

The volume of the oxygen tubes used was approximately 200 cubic millimeters. The area of the cathode surface of the cell was 8.4 sq. cm and the total area of the cathode surface and the shielding cylinder covered by the active material in sputtering was approximately 43 sq. cm.

THORIUM, URANIUM AND CALCIUM

The relative sensitivity for equal energy of different wave-lengths of a pure thorium cell in

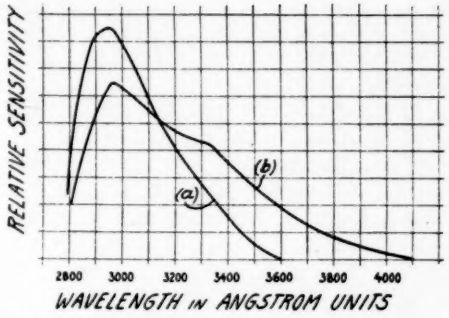


FIG. 2. (a) Sensitivity curve for a thorium cell in Corex D glass. (b) Sensitivity curve of the same cell after reaction with the oxygen from a tube at 20 mm pressure.

Corex D glass is shown by curve (a) Fig. 2. When an oxygen tube at about 20 mm pressure is broken in this cell the active surface reacts with the oxygen as is indicated by the change in the photoelectric response. The cell after the oxygen has reacted has the current-voltage characteristic of a vacuum cell and the response is not altered by re-exhausting as previously described. The sensitivity for the new surface is shown by curve (b) Fig. 2. The change in sensitivity clearly depends upon the area of the active surface and the amount of oxygen introduced. In all cases where small amounts of oxygen were introduced the result was a shift in the photoelectric threshold to the same longer wave-length while an excess of oxygen results in the formation of the normal oxide of thorium and the photoelectric activity is again completely destroyed except for the very short radiations to which the normal oxide is

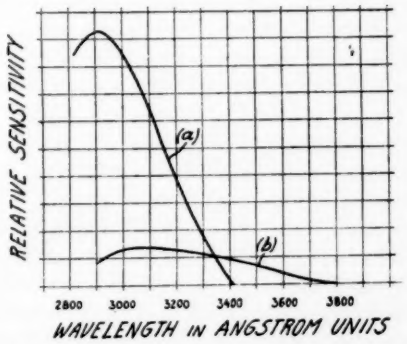


FIG. 3. (a) Sensitivity curve for a uranium cell in Corex D glass. (b) Sensitivity curve of the same cell after reaction with oxygen from a 40-mm pressure tube.

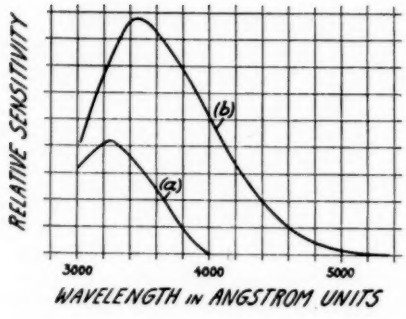


FIG. 4. (a) Sensitivity curve for a calcium cell in Corex D glass. (b) Sensitivity curve for the same cell treated with the oxygen from a tube at 40 mm pressure.

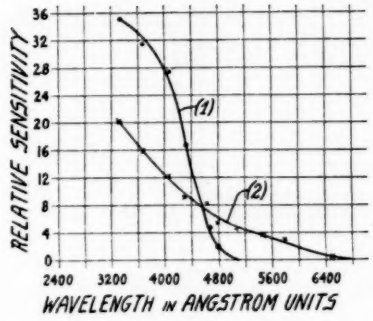


FIG. 5. (1) Sensitivity curve for barium metal in Corex D glass. (2) Sensitivity curve for the vaporized deposit from a barium oxide coated cathode.

sensitive. Figs. 3 and 4 show a similar effect of oxygen upon uranium and calcium as is obtained with thorium. Sensitivity of the oxygen-treated calcium surface was unmistakably detectable for the green mercury radiation of wave-length 5460A.

BARIUM

The response curves for barium (Fig. 5) and for the active material obtained from an activated oxide coated surface were previously obtained.[1] The threshold wave-length for barium metal as indicated from this curve is approximately 5200A. If the sensitivity of the measuring device is increased a hundred-fold or more, a very faint response to the green and yellow mercury lines is still barely detectable. This is probably due to a slight contamination of the surface with the sensitive constituent found in activated oxide coated cathodes. This view is further verified by the fact that as more and more oxygen in these small doses is added up to a certain limit to a photo-cell of barium the response to the yellow and green radiation from mercury is materially increased while the sensitivity to the shorter wave-lengths (for the blue, violet, and ultraviolet), is decreased.

For example, to the same cell were attached a number of side tubes so that first an oxygen tube at 20 mm pressure was introduced into the cell. This was allowed to react and then the cell was re-exhausted and again tested. After this an oxygen tube at 40 mm pressure was broken and the response noted after re-exhausting. Finally 80-mm pressure and 160-mm pressure oxygen tubes were introduced. The response to the yellow mercury lines was increased 40-fold while the sensitivity to the green mercury radiation at 5460A was increased 60-fold. At the same time the blue, violet and ultraviolet response were each diminished to less than one-half the initial value. Further, the original cell showed no response that could be detected from the radiation from a 500-watt Mazda lamp through a filter, transparent only to 6900A and longer radiation, while the treated surfaces show a decided response to this light. An excess of oxygen again forms the normal oxide and destroys the activity except for very short radiation.

CAESIUM

A caesium photoelectric cell with the caesium distilled into the cell and deposited on magnesium was made and tested for light passing through different filters. Oxygen tubes at 20 and 40 mm pressure were introduced into the cell. The cell was tested at once, then re-exhausted and again tested. The total response to a mazda lamp before and after the re-exhaust was increased only a few percent while the response to light through a deep red filter transmitting only wave-lengths longer than 6900A was increased by more than one thousand percent. A 300-cubic millimeter tube of oxygen under a pressure of 450 mm again destroyed all activity.

To take into account the possible effect of caesium which might vaporize and again deposit on the surface a second cell was made as above.

The introduction of oxygen again increased the relative sensitivity in the red. This cell was not re-exhausted and in two months has not changed. If free oxygen was still present at the time of the original test the sensitivity should have decreased while an excess of free caesium due to the volatility of caesium should now cover the surface with a layer of more than atomic thickness and produce only normal caesium activity.

A caesium cell was next made by first depositing silver on glass. Caesium was then distilled onto the silver and the response through different filters noted. Oxygen was then admitted in small doses. Again the red sensitivity increased rapidly, while the total response to white light, rose only slightly. The red sensitivity is unmistakably present even before all the oxygen is used up. Oxygen was then admitted until the caesium layer was completely oxidized and the activity destroyed. Caesium was now allowed to slowly diffuse into the cell somewhat similar to the method described by Campbell[4] and Koller[5] except that the cell proper was kept at room temperature. The progressive change in photosensitivity was noted both for white light and for light passing through a red filter. The ratio of the response to light through a red glass filter to the total response without the filter shows that a decided red sensitivity remained practically constant throughout the test even though sufficient caesium was introduced to coat the window, indicating a definite reaction between the caesium and caesium oxide on the silver.

Zirconium, Titanium, Iron and Nickel

Similar tests for the effect of oxygen upon zirconium, titanium, iron and nickel show that when oxygen in ever so small an amount is introduced the effect is always to reduce the photoelectric emission and shift the threshold toward the shorter wave-length.

Silver

The effect of oxygen upon the photoelectric response of silver is the same as with zirconium and titanium, but the original sensitivity is readily restored by again heating by high frequency the active surface with the cell on a high vacuum pump. This is to be expected since silver oxide is easily reduced by heat.

Gold

The threshold wave-length of gold is somewhat longer than for platinum. All tests on this metal were made through various filters with a continuous hydrogen discharge tube as a light source. No response could be detected through 2 cm clear fused quartz. An absorption cell with crystal quartz plates each about 2 mm thick reduces the response to approximately one-half while the same cell (16 mm between plates). filled with distilled water, only very slightly decreased the response below that through the empty cell. A solution of four drops of acetic acid in 30 cc distilled water in the same cell reduced the response to about 3 percent of that through the empty cell. A 200-cubic millimeter tube of oxygen at 10 mm pressure was broken. No voltage was applied to the cell. Several hours afterward, this oxygen was again pumped out and the cell tested. There was no detectable change in the response. A second oxygen tube at 20 mm pressure was next broken. The cell was exposed for about five hours to the same light source as was used in testing with 135 volts across the cell. The cell was again re-exhausted and still no change in response could be detected.

Conclusions

The change in activity by the action of oxygen with such metals as titanium, silver, etc., may be explained as due to the formation of the normal oxides. This is further indicated by the ease with which the original activity of silver is again restored by heating and the stability of a gold cell. Thorium, uranium, etc. (except caesium) all have high melting points and too low vapor pressures at the temperatures at which the oxygen was introduced into the cells, and the amounts of oxygen added were too great to explain these threshold shifts as due to the formation of a manotomic layer of the metal on the surface as suggested by Koller[5] and Campbell[4] for caesium. The increased sensitivity and longer threshold wave-length for these metals appears to be due to the activity of the product formed by

[4] Campbell, Phil. Mag. **6**, 633 (1928).
[5] Koller, Phys. Rev. **36**, 1639 (1930).

the interaction of the oxygen with the metal (Case[6] suggests a suboxide to explain the activity with barium.)

The photoelectric work functions for the oxygen-treated surfaces of thorium, uranium and calcium, namely, 3.01, 3.25 and 2.26 volts, respectively, calculated from the photoelectric threshold wave-lengths here obtained, agree remarkably well with the thermionic work functions 2.94, 3.28 and 2.24 volts obtained by Dushman[2] for activated tungsten filaments containing oxides of these same metals. This suggests that possibly the thermionic emission for these activated tungsten filaments may be due to a layer of the same active material as that from which the increased photoelectric effects were obtained but does not prove anything as regards the thickness of the layer.

[6] Case, *Photo-Electric Cells and Their Applications* (The Physical and Optical Societies, London, 1930), p. 54.

A Simplified Theory of the Image Errors According to L. Seidel*

M. HERZBERGER, *Eastman Kodak Company*

(Received October 19, 1935)

DURING the past century the theory of L. Seidel was a powerful tool in the hand of the optical designer. Moser, Schwarzschild, Kerber, and others have shown how this method was used especially in the theory of astronomical objectives and eyepieces. Since these methods, however, did not hold for systems with higher aperture, because the image errors of higher orders play an essential part, the designers in most countries went back to the old method of tracing rays. In 1930, M. Berek,[1] applied the knowledge of the Seidel coefficients to systems of large aperture. Here, the coefficients indicate the amount each surface gives to the single image error, and which surfaces are effective for altering these amounts; Berek's method shows, however, that the building up of optical systems with the analytical apparatus of Seidel is rather complicated.

The author believes, therefore, that a simplification of the formulae will be welcome to all working in the field of optical design. Theoretical investigations show that the center of a spherical surface is more important than the vertex. The author therefore presents in the first section of the present paper a new method of tracing paraxial and finite rays through an optical system. The method is shorter, especially for machine computation, than the method usually used, and has the theoretical advantage that it deals only with optical lengths of paths.

The author develops his method from the designing formulae usually used, and indicates the connection of the new image error formulae with the old ones.

He shows how to obtain the Gaussian data for an arbitrary ray, if a single one is traced. For completeness, he discusses the single errors and gives the variation according to a change of stop and object. These formulae, published here for the first time, make it possible to investigate the image errors of every object, if an arbitrary paraxial ray has been computed through the system. Now the way is clear to investigate systematically, for instance, the characteristics of a single lens and to search for a system which gives a perfect image. Such systems must have vanishing Seidel aberrations.[2] Besides the usual image errors, the author also deals in this paper with the image errors for curved object and image.

At the end is given a classification of the systems with a high degree of freedom from image errors, and a description of their characteristics. These investigations lead to an interesting simple law, which must be fulfilled in systems with sharp image formation. Concentric systems and systems having at most three surfaces are also considered.

In developing the subject, a carefully chosen nomenclature and notation is a valuable aid. We suggest that:

1. All optical lengths (length multiplied by its refractive index) be labeled with the same letter as the length, but in bold-face type.
2. Lengths be written with small Roman letters, and their reciprocals with the same letters capitalized.
3. A magnitude after refraction have the same letter as before, but with a prime.
4. The variation of a magnitude with refraction be indicated by placing the symbol Δ before the magnitude.
5. All lengths be denoted by Roman letters, and all magnitudes without dimension by Greek letters.

These conventions are mostly in accord with those given in the German Normblatt, DIN 1335 (Bezeichnungen in der technischen Optik).

Especially we denote the distance from object to pole by "s," and from object to center by "c."

Some examples:

$$\mathbf{c}=nc, \quad C=1/c, \quad \Delta\mathbf{C}=1/n'c'-1/nc.$$

* Communication No. 558 from the Kodak Research Laboratories.

[1] M. Berek, *Grundlagen der Praktische Optik* (Verlag W. de Gruyter, Berlin and Leipzig, 1930).

[2] Until now the only known perfect systems were the single spherical surface (for its aplanatic points), and the systems given by H. Boegehold and the author ["Kugelsymmetrische Systeme," Zeits. f. ang. Math. **15**, 157–58 (1935)].

Section I. Designing Formulae with the Center as Origin

We begin with the known paraxial formula:

$$\Delta nS = R\Delta n, \tag{1, 1}$$

in which n and n' are the refractive indices of the media separated by the refracting surface, the curvature of which is given by R.

We note the identities:

$$s = c + r, \quad c = s - r. \tag{1, 2}$$

Instead of (1) we can write:

$$\Delta n(S - R) = 0$$

or $$n(s-r)/sr = n'(s'-r)/s'r$$

or $$nc/s = n'c'/s'. \tag{1, 3}$$

From (3) we get

$$(c+r)/ncr = (c'+r)/n'c'r$$

or $$\boxed{\Delta(\mathbf{C}+\mathbf{R}) = 0} \tag{1, 4}$$

or $$\Delta\mathbf{C} = -\Delta\mathbf{R}.$$

Attention is called to the fact that on the right side is the so-called Petzval sum of a single surface, an expression which is needed and used in all calculating work.

The transposition formulae are as simple as usual. If we have the magnitude

$$\mathbf{m}_{1,2} = n_{1,2}(r_2 - r_1 + d_{1,2}) \tag{1, 5}$$

where $d_{1,2}$ is the distance between the first two poles, and $\mathbf{m}_{1,2}$ the optical distance between the first two centers, we see that

$$\boxed{\mathbf{c}_2 = \mathbf{c}_1' - \mathbf{m}_{1,2}}. \tag{1, 6}$$

Eqs. (4) and (6) give the Gaussian calculation. If we write the result in closed form, we have a continuous fraction:

$$\mathbf{c}_K' = \cfrac{1}{-\Delta_K\mathbf{R} + \cfrac{1}{-\mathbf{m}_{K,K-1} + \cfrac{1}{-\Delta_{K-1}\mathbf{R} + \cfrac{1}{-\mathbf{m}_{K-1,K-2} + \cfrac{1}{\cdots + \cfrac{1}{\cdots + \cfrac{1}{-\mathbf{m}_{1,2} + \cfrac{1}{-\Delta_1\mathbf{R} + \cfrac{1}{\mathbf{c}}}}}}}}}}. \tag{1, 7}$$

It is easily seen how well this method is suited for machine calculation.

The angle magnification $\gamma' = \sigma'/\sigma$, (σ, σ' = angles of a ray with the axis), is usually obtained by the formula:

$$\gamma' = \Pi\frac{s}{s'} = \frac{s_1 \cdot s_2 \cdot s_3 \cdots s_K}{s_1' \cdot s_2' \cdot s_3' \cdots s_K'}. \tag{1, 8}$$

The letter Π indicates as usual the product of the magnitude after it, taken over all the surfaces.

We find from (3)

$$s/s' = \mathbf{c}/\mathbf{c}' = \mathbf{C}'/\mathbf{C}; \tag{1, 9}$$

and from that

$$\gamma' = \Pi\mathbf{c}/\mathbf{c}' = \Pi\mathbf{C}'/\mathbf{C}. \tag{1, 10}$$

Instead of the linear magnification β', we use, according to the principle of using optical lengths, the reduced magnification: $\mathfrak{B}' = (n'/n)\beta'$ where n', n are the indices of the image and object spaces.

The law of Lagrange shows

$$\mathfrak{B}'\gamma' = 1; \tag{1, 11}$$

and that means

$$\mathfrak{B}' = \Pi \mathbf{c}'/\mathbf{c} = \Pi \mathbf{C}/\mathbf{C}'. \qquad (1, 12)$$

We are accustomed to compute the magnification in the following way: The height h at which a paraxial ray meets a surface is given by the equation:

$$h = s\sigma = s'\sigma'. \qquad (1, 13)$$

Here σ means the angle between the paraxial ray and the axis. We have further

$$h_\nu/h_{\nu-1} = s_\nu/s_{\nu-1}. \qquad (1, 14)$$

We calculate h_ν/h_1 and get, by means of formula (1, 8):

$$\beta' = \frac{n}{n'\gamma'} = \frac{ns_\nu'}{n's_1} \Big/ \frac{h_\nu}{h_1}. \qquad (1, 15)$$

The expressions h_ν/h_1, which we get here, have their own importance, and are especially valuable in the Seidel theory.

We can get the corresponding expressions if we put in

$$\mathbf{p} = \mathbf{c}\sigma = \mathbf{c}'\sigma' \qquad (1, 16)$$

which follows from (9) and (13). Then

$$\mathbf{p}_\nu/\mathbf{p}_{\nu-1} = \mathbf{c}_\nu/\mathbf{c}'_{\nu-1} = \mathbf{C}'_{\nu-1}/\mathbf{C}_\nu \qquad (1, 17)$$

and

$$\mathfrak{B}' = \frac{\sigma_1}{\sigma_\nu'} = \frac{\mathbf{c}_\nu'}{\mathbf{c}_1} \Big/ \frac{\mathbf{p}_\nu}{\mathbf{p}_1} = \frac{\mathbf{C}_1}{\mathbf{C}_\nu'} \Big/ \frac{\mathbf{p}_\nu}{\mathbf{p}_1}. \qquad (1, 18)$$

But the geometrical meaning of p is different. Eqs. (1, 16) shows that p is the perpendicular distance from the center to the ray.

The special case of a parallel entering ray, $C_1 = 0$, obviously gives the power Φ of the system:

$$\Phi = \mathbf{C}_\nu' \cdot \mathbf{p}_\nu/\mathbf{p}_1 = \sigma_\nu'/\mathbf{p}_1. \qquad (1, 19)$$

We note

$$\Phi = \lim_{C_1 \to 0} \frac{\mathbf{C}_1}{\mathfrak{B}'}. \qquad (1, 20)$$

Formulae (4, 6, 14, 18, 19) are those needed in paraxial calculation. To complete the system, we add the formulae for finite rays. These are nearly the same as the usual forms, except for the abbreviation of using only optical lengths. The fundamental formulae are

$$\mathbf{p} = \mathbf{c} \sin \sigma = \mathbf{r} \sin \epsilon = \mathbf{r}' \sin \epsilon' = \mathbf{c}' \sin \sigma', \qquad (1, 21)^3$$
$$\epsilon + \sigma = \epsilon' + \sigma', \quad \mathbf{c}_2 = \mathbf{c}_1' - \mathbf{m}_{1,2}$$

where ϵ, ϵ' are angles of incidence and refraction.

Section II. Two Formulae Corresponding to Those Found by L. Seidel, and Their Application

The purpose of this section is to show how to get the data of an arbitrary ray in the simplest manner after having traced a single one. The solution of this problem corresponds to that given by Berek for the more complicated vertex formulae. The principal formulae which we derive correspond to some given by Seidel, who, however, used them only to develop the image errors in relation to the stop.

A magnitude which does not change by refraction is called a surface invariant. According to (1, 4) and (1, 16), these expressions are such invariants:

$$I = \mathbf{C} + \mathbf{R} \quad \text{and} \quad \mathbf{p}$$

which means

$$\Delta I = \Delta(\mathbf{C} + \mathbf{R}) = 0, \quad \Delta \mathbf{p} = 0. \qquad (2, 1)$$

We have here invariants for refraction: What we want to find is an expression which remains invariant for the whole system. For this purpose we need two rays. Let us denote all quantities pertaining to the second ray by the subscript D, standing for "diaphragm," because we shall mostly choose the second ray such that it passes through the center of the exit pupil. Then, obviously

$$\mathbf{c}_{\nu+1} - \mathbf{c}_{\nu+1,D} = \mathbf{c}'_\nu - \mathbf{c}'_{\nu,D}. \qquad (2, 2)$$

By applying (1, 16)

$$\frac{\mathbf{p}_{\nu+1}}{\sigma_{\nu,\nu+1}} - \frac{\mathbf{p}_{\nu+1,D}}{\sigma_{\nu,\nu+1,D}} = \frac{\mathbf{p}_\nu}{\sigma_{\nu,\nu+1}} - \frac{\mathbf{p}_{\nu,D}}{\sigma_{\nu,\nu+1,D}}$$

or

$$\mathbf{p}_{\nu+1}\sigma_{\nu,\nu+1,D} - \mathbf{p}_{\nu+1,D}\sigma_{\nu,\nu+1} = \mathbf{p}_\nu\sigma_{\nu,\nu+1,D} - \mathbf{p}_{\nu,D}\sigma_{\nu,\nu+1}$$

[3] Compare Southall, J. Opt. Soc. Am. **5**, 19 (1921).

or, again applying (1, 16),

$$\mathfrak{p}_{\nu+1}\mathfrak{p}_{\nu+1,\,D}(\mathfrak{C}_{\nu+1}-\mathfrak{C}_{\nu+1,\,D}) = \mathfrak{p}_{\nu}\mathfrak{p}_{\nu,\,D}(\mathfrak{C}_{\nu}-\mathfrak{C}_{\nu,\,D}). \quad (2,\,3)$$

Now we see that

$$\mathfrak{C}_{\nu+1}-\mathfrak{C}_{\nu+1,\,D}=I_{\nu+1}-I_{\nu+1,\,D}. \quad (2,\,4)$$

We notice that all magnitudes in (3) are invariant for refraction, and from that comes the important formula:

$$\mathfrak{p}_{\nu}\mathfrak{p}_{\nu,\,D}(I_{\nu}-I_{\nu,\,D})=\mathfrak{p}_1\mathfrak{p}_{1,\,D}(I_1-I_{1,\,D}) \quad (2,\,5)$$

or
$$\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\left(\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\right)_D(I_{\nu}-I_{\nu,\,D})=I_1-I_{1,\,D}.$$

Now, if we could find expressions which allow us to develop $(\mathfrak{p}_{\nu}/\mathfrak{p}_1)_D$ with the help of the data found for the first ray, (5) would enable us to get the data for every object point by tracing only a single ray through the system.

First, we introduce for abbreviation

$$I_{\nu}-I_{\nu,\,D}=\mathfrak{C}_{\nu}-\mathfrak{C}_{\nu,\,D}=D_{\nu} \quad (2,\,6)$$

and write (2, 5)

$$\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\left(\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\right)_D D_{\nu}=D_1. \quad (2,\,5a)$$

Formula (1, 17) shows

$$\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_{\nu-1}}=\frac{\mathfrak{c}_{\nu}}{\mathfrak{c}'_{\nu-1}}=1-\mathfrak{m}_{\nu,\,\nu-1}\mathfrak{C}'_{\nu-1}. \quad (2,\,7)$$

Therefore

$$\left(\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_{\nu-1}}\right)_D-\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_{\nu-1}}=\mathfrak{m}_{\nu,\,\nu-1}D_{\nu-1} = \frac{\mathfrak{m}_{\nu,\,\nu-1}}{(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)_D}\cdot D_1 \quad (2,\,8)$$

and from (8) we derive

$$\frac{\mathfrak{p}_{\nu,\,D}}{\mathfrak{p}_{\nu}}-\frac{\mathfrak{p}_{\nu-1,\,D}}{\mathfrak{p}_{\nu-1}}=\frac{\mathfrak{p}_{1,\,D}}{\mathfrak{p}_1}\cdot\frac{\mathfrak{m}_{\nu,\,\nu-1}}{(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)\cdot(\mathfrak{p}_{\nu}/\mathfrak{p}_1)}\cdot D_1, \quad (2,\,9)$$

which gives

$$\frac{\mathfrak{p}_{\nu,\,D}}{\mathfrak{p}_{\nu}}=\frac{\mathfrak{p}_{1,\,D}}{\mathfrak{p}_1}\left[1+D_1\Sigma\frac{\mathfrak{m}_{\nu,\,\nu-1}}{(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)\cdot(\mathfrak{p}_{\nu}/\mathfrak{p}_1)}\right]$$

or

$$\left(\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\right)_D=\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\left[1+D_1\Sigma\frac{\mathfrak{m}_{\nu,\,\nu-1}}{(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)\cdot(\mathfrak{p}_{\nu}/\mathfrak{p}_1)}\right]. \quad (2,\,10)$$

This formula (10) solves the question we put. We abbreviate

$$\Sigma\frac{\mathfrak{m}_{\nu,\,\nu-1}}{(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)\cdot(\mathfrak{p}_{\nu}/\mathfrak{p}_1)}=Z_{\nu} \quad (2,\,11)$$

and write (10) in this form:

$$\boxed{\left(\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}\right)_D=\frac{\mathfrak{p}_{\nu}}{\mathfrak{p}_1}[1+D_1Z_{\nu}]}. \quad (2,\,12)$$

Now we can get the data of any ray from D_1 and the data of the first ray. We find from (2, 5)

$$\mathfrak{C}'_{\nu,\,D}=\mathfrak{C}_{\nu}'-D_{\nu}=\mathfrak{C}_{\nu}'-\frac{D_1}{(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+D_1Z_{\nu})}. \quad (2,\,13)$$

From (1, 18) we find

$$\beta_D'=\frac{\mathfrak{C}_{1,\,D}}{\mathfrak{C}'_{\nu,\,D}(\mathfrak{p}_{\nu}/\mathfrak{p}_1)_D} =\frac{(\mathfrak{C}_1-D_1)(\mathfrak{p}_{\nu}/\mathfrak{p}_1)}{\mathfrak{C}_{\nu}'(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+D_1Z_{\nu})-D_1} \quad (2,\,14)$$

and (13) gives finally

$$\mathfrak{c}'_{\nu,\,D}=\frac{(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+D_1Z_{\nu})}{\mathfrak{C}_{\nu}'(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+D_1Z_{\nu})-D_1}. \quad (2,\,15)$$

Eqs. (2, 5) and (2, 10) correspond to the formulae given by Seidel for the pole-to-pole calculation. Eqs. (13) and (15) correspond to the formulae given by M. Berek.[4]

In particular, we find the distance of the focal point by putting $\mathfrak{C}_{1,\,D}=0$; $D_1=\mathfrak{C}_1$ which gives

$$\mathfrak{c}'_{\nu,\,\infty}=\frac{(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+\mathfrak{C}_1Z_{\nu})}{\mathfrak{C}_{\nu}'(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+\mathfrak{C}_1Z_{\nu})-\mathfrak{C}_1} \quad (2,\,16)$$

and the power according to (1, 20) by

$$\Phi=\frac{\mathfrak{C}_{\nu}'(\mathfrak{p}_{\nu}/\mathfrak{p}_1)^2(1+\mathfrak{C}_1Z_{\nu})-\mathfrak{C}_1}{\mathfrak{p}_{\nu}/\mathfrak{p}_1}. \quad (2,\,17)$$

[4] M. Berek, Cent. Ztg. f. Opt. u. Mech. **44**, 103 (1923).

Now we will find the position of the nodal points. To do this, we must make $\beta'=1$, and eliminate D_1. We get finally

$$\mathfrak{c}_{1,\,K}=\frac{(\mathfrak{p}_\nu/\mathfrak{p}_1)((\mathfrak{p}_\nu/\mathfrak{p}_1)\mathfrak{C}_\nu'Z_\nu+1)-1}{(\mathfrak{p}_\nu/\mathfrak{p}_1)^2\mathfrak{C}_\nu'(\mathfrak{C}_1Z_\nu+1)-\mathfrak{C}_1}, \quad (2,\,18)$$

$$\mathfrak{c}'_{\nu,\,K}=\frac{(\mathfrak{p}_\nu/\mathfrak{p}_1)(1+\mathfrak{C}_1Z_\nu)-1}{(\mathfrak{p}_\nu/\mathfrak{p}_1)^2\mathfrak{C}_\nu'(1+\mathfrak{C}_1Z_\nu)-\mathfrak{C}_1}.$$

In the same way we can find the data for the ray in the object space if the data in the image space are given. We get

$$D_1=\frac{(\mathfrak{p}_\nu/\mathfrak{p}_1)^2D_\nu}{1-(\mathfrak{p}_\nu/\mathfrak{p}_1)^2D_\nu Z_\nu},$$

$$\left(\frac{\mathfrak{p}_\nu}{\mathfrak{p}_1}\right)_D=\frac{\mathfrak{p}_\nu}{\mathfrak{p}_1}\cdot\frac{1}{1-(\mathfrak{p}_\nu/\mathfrak{p}_1)^2D_\nu Z_\nu}, \quad (2,\,19)$$

$$\beta'=\frac{\mathfrak{C}_1(1-(\mathfrak{p}_\nu/\mathfrak{p}_1)^2D_\nu Z_\nu)-(\mathfrak{p}_\nu/\mathfrak{p}_1)^2D_\nu}{(\mathfrak{C}_\nu'-D_\nu)(\mathfrak{p}_\nu/\mathfrak{p}_1)}.$$

As a special case, we show the expression for the focus in the object space:

$$\mathfrak{c}_{1,\,\infty}=\frac{1-(\mathfrak{p}_\nu/\mathfrak{p}_1)^2\mathfrak{C}_\nu'Z_\nu}{\mathfrak{C}_1-(\mathfrak{p}_\nu/\mathfrak{p}_1)^2\mathfrak{C}_\nu'(1+Z_\nu\mathfrak{C}_1)} \quad (2,\,20)$$

and for the power

$$\Phi'=-\frac{(\mathfrak{p}_\nu/\mathfrak{p}_1)^2\mathfrak{C}_\nu'(1+Z_\nu\mathfrak{C}_1)-\mathfrak{C}_1}{\mathfrak{p}_\nu/\mathfrak{p}_1} \quad (2,\,21)$$

in agreement with (2, 17).

We can also proceed in the opposite direction. If we write for abbreviation

$$Z_{\nu,\,D}=\sum\frac{\mathfrak{m}_{\nu,\,\nu-1}}{(\mathfrak{p}_\nu/\mathfrak{p}_1)_D(\mathfrak{p}_{\nu-1}/\mathfrak{p}_1)_D} \quad (2,\,22)$$

we have

$$\mathfrak{p}_\nu/\mathfrak{p}_1=(\mathfrak{p}_\nu/\mathfrak{p}_1)_D(1-D_1Z_{\nu,\,D}) \quad (2,\,23)$$

and

$$D_\nu=\frac{D_1}{(\mathfrak{p}_\nu/\mathfrak{p}_1)_D^2(1-D_1Z_{\nu,\,D})} \quad (2,\,24)$$

Z_ν and $Z_{\nu,\,D}$ are related by

$$(1+D_1Z_\nu)(1-D_1Z_{\nu,\,D})=1 \quad (2,\,25)$$

or:

$$Z_{\nu,\,D}=Z_\nu/(1+D_1Z_\nu),\qquad Z_\nu=Z_{\nu,\,D}/(1-D_1Z_{\nu,\,D}). \quad (2,\,26)$$

Section III. Some Relations Between Expressions in Gaussian Optics. Some Important Constants

We repeat the Gaussian formulae:

$$\begin{aligned}&\mathfrak{A}(\mathfrak{C}+\mathfrak{R})=\mathfrak{A}I=0,\\ &I-I_D=\mathfrak{C}-\mathfrak{C}_D=\mathfrak{C}'-\mathfrak{C}_D'=D,\\ &\mathfrak{A}\mathfrak{p}=0,\quad \mathfrak{p}\mathfrak{C}=\sigma.\end{aligned} \quad (3,\,1)$$

It is obvious that an invariant standing as a factor after the sign $\mathfrak{A}$ may be placed before the sign. With the help of this theorem, we find, for instance

$$\begin{aligned}&\mathfrak{A}\mathfrak{C}\mathfrak{R}=\mathfrak{A}(\mathfrak{C}_D+D)\mathfrak{R}=\mathfrak{A}\mathfrak{C}_D\mathfrak{R}+D\mathfrak{A}\mathfrak{R},\\ &I=I_D+D,\\ &\mathfrak{A}\mathfrak{C}^2=\mathfrak{A}\mathfrak{C}(I-\mathfrak{R})=-(I\mathfrak{A}\mathfrak{R}+\mathfrak{A}\mathfrak{C}\mathfrak{R}),\\ &\mathfrak{A}\mathfrak{C}\mathfrak{C}_D=\mathfrak{A}\mathfrak{C}(I_D-\mathfrak{R})=\mathfrak{A}\mathfrak{C}_D(I-\mathfrak{R})\\ &\qquad=-(I\mathfrak{A}\mathfrak{R}+\mathfrak{A}\mathfrak{C}_D\mathfrak{R})=-(I_D\mathfrak{A}\mathfrak{R}+\mathfrak{A}\mathfrak{C}\mathfrak{R}),\\ &\mathfrak{A}\mathfrak{C}_D^2=-(I_D\mathfrak{A}\mathfrak{R}+\mathfrak{A}\mathfrak{C}_D\mathfrak{R})\\ &\qquad=-(I\mathfrak{A}\mathfrak{R}+\mathfrak{A}\mathfrak{C}\mathfrak{R}-2D\mathfrak{A}\mathfrak{R}).\end{aligned} \quad (3,\,2)$$

We shall now develop some other formulae. These equations are valid and serve to define ρ, σ, and τ.

$$\rho=\sum(\mathfrak{p}_\nu/\mathfrak{p}_1)^2\mathfrak{A}\mathfrak{C}^2=(1/\mathfrak{p}_1^2)\sum\mathfrak{A}\sigma^2$$
$$=(1/\mathfrak{p}_1^2)\sum(\sigma'^2-\sigma^2)=\mathfrak{C}_1^2(1/\beta^2-1), \quad (3,\,3)$$

$$\sigma=\sum(\mathfrak{p}_\nu/\mathfrak{p}_1)(\mathfrak{p}_\nu/\mathfrak{p}_1)_D\mathfrak{A}\mathfrak{C}\mathfrak{C}_D$$
$$=\mathfrak{C}_1\mathfrak{C}_{1,\,D}(1/\beta'\beta_D'-1), \quad (3,\,4)$$

$$\tau=\sum(\mathfrak{p}_\nu/\mathfrak{p}_1)_D^2\mathfrak{A}\mathfrak{C}_D^2=\mathfrak{C}^2{}_{1,\,D}(1/\beta_D''-1). \quad (3,\,5)$$

These three constants will play an important role in our further investigations. We shall discuss them.

First, we see that ρ can vanish if and only if the object point coincides with one of the nodal points. $\beta'=\pm1$.

σ can vanish only if the object and the stop have a *reciprocal reduced magnification*. Two points which have *reciprocal reduced magnification* we shall call *reciprocal points*. Then we

can say that σ vanishes only if object and stop are reciprocal points.

τ can vanish only if the stop is at one of the nodal points.

The three constants, ρ, σ, τ, can vanish simultaneously only when both object and stop are situated at a positive or negative nodal point. In this case, the system must be telescopic with reduced magnification plus or minus one. Then, every point on the axis of the system is a nodal point. For abbreviation, we denote such a system as a *nodal system.*

From (1, 20) we find the limits of our three constants, if object or stop moves to infinity.

We can develop ρ, σ, τ with reference to the object or to the stop. If we consider them

$$\begin{aligned} \rho &= \rho_0 &&= \tau_D - 2D_1\sigma_D + D_1^2\rho_D, \\ \sigma &= \rho_0 + D_1\sigma_0 &&= \tau_D - D_1\sigma_D, \\ \tau &= \rho_0 + 2D_1\sigma_0 + D_1^2\tau_0 = \tau_D \end{aligned} \tag{3, 6}$$

with reference to object. with reference to stop.

From this we have:

$$\begin{aligned} \rho_0 &= \sum (\mathbf{p}_\nu/\mathbf{p}_1)^2 \Delta\mathbf{C}^2, \\ \rho_D &= \sum (\mathbf{p}_\nu/\mathbf{p}_1)_D^2 Z_{\nu,D}^2 \Delta\mathbf{C}_D^2 + 2Z_{\nu,D}\Delta\mathbf{R}, \\ \sigma_0 &= \sum (\mathbf{p}_\nu/\mathbf{p}_1)^2 Z_\nu \Delta\mathbf{C}^2 + \Delta\mathbf{R}, \\ \sigma_D &= \sum (\mathbf{p}_\nu/\mathbf{p}_1)_D^2 Z_{\nu,D} \Delta\mathbf{C}_D^2 + \Delta\mathbf{R}, \\ \tau_0 &= \sum (\mathbf{p}_\nu/\mathbf{p}_1)^2 Z_\nu^2 \Delta\mathbf{C}^2 + 2Z_\nu\Delta\mathbf{R}, \\ \tau_D &= \sum (\mathbf{p}_\nu/\mathbf{p}_1)_D^2 \Delta\mathbf{C}_D^2. \end{aligned} \tag{3, 7}$$

By resolving (3, 6) we find

$$\begin{aligned} \rho_0 &= \rho, & \rho_D &= \frac{1}{D_1^2}(\rho + \tau - 2\sigma), \\ \sigma_0 &= \frac{1}{D_1}(\sigma - \rho), & \sigma_D &= \frac{1}{D_1}(\tau - \sigma), \\ \tau_0 &= \frac{1}{D_1^2}(\rho + \tau - 2\sigma), & \tau_D &= \tau; \end{aligned} \tag{3, 8}$$

$\rho_D = \tau_0$. Therefore ρ_D and τ_0 must be independent of the object coordinates as well as of the coordinates of the ray in the plane of the stop.

If we combine (3, 5) with (3, 8), we get

$$\begin{aligned} \rho_0 &= \mathbf{C}_1^2(1/\beta'^2 - 1), \\ \rho_D &= (1/\beta_0'^2 - 1), \\ \sigma_0 &= -\mathbf{C}_1(1/\beta'\beta_0' + 1), \\ \sigma_D &= -\mathbf{C}_{1,D}(1/\beta_0'\beta_D' + 1), \\ \tau_0 &= (1/\beta_0'^2 - 1), \\ \tau_D &= \mathbf{C}^2{}_{1,D}(1/\beta_D'^2 - 1). \end{aligned} \tag{3, 9}$$

Here we have put for abbreviation

$$\frac{1}{\beta_0'} = \frac{\mathbf{C}_{1,D}/\beta_D' - \mathbf{C}_1/\beta'}{\mathbf{C}_{1,D} - \mathbf{C}_1}. \tag{3, 10}$$

The geometrical meaning of β_0' may be seen if we let the object or the stop shift to the center of the first surface. (3, 10), which must remain independent of the shift, assumes the value $\beta_0' \to \beta'$ or β_D', respectively. This means that β_0' is the reduced magnification of an object at the center of the first surface.

Section IV. Transformation of the Aberration Formula of L. Seidel

To define the Seidel aberrations in the manner in which they will be used here, we consider Eq. (4, 2).

A ray may be defined by the optical coordinates (multiplied by its refractive index) of its intersection point in the object plane, and the optical coordinates of its intersection point in the exit pupil. These are denoted, respectively, by: $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x}_D'$, $\mathbf{y}_D'$.

For abbreviation

$$\begin{aligned} \mathbf{a} &= \mathbf{x}^2 + \mathbf{y}^2 = n^2(x^2 + y^2), \\ \mathbf{b} &= 2(\mathbf{x}\mathbf{x}_D' + \mathbf{y}\mathbf{y}_D') = 2nn'(x x_D' + y y_D'), \\ \mathbf{c} &= \mathbf{x}_D'^2 + \mathbf{y}_D'^2 = n'^2(x_D'^2 + y_D'^2). \end{aligned} \tag{4, 1}$$

We shall let $\mathbf{x}_0'$, $\mathbf{y}_0'$ be the coordinates of the point at which the ray intersects the plane through the Gaussian image point. If the object has an exact image, the expressions $\mathbf{x}_0' - \beta'\mathbf{x}$, $\mathbf{y}_0' - \beta'\mathbf{y}$ must be zero for all rays. If the image is not perfect, we can develop these expressions in relation to $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x}_D'$, $\mathbf{y}_D'$. If we consider the axial symmetry of the system, and standardize the coefficients[5] in a convenient manner, we have

[5] See M. Herzberger, *Strahlenoptik* (Verlag J. Springer, Berlin, 1931), p. 116, formula 14.

$$\mathbf{x}_0'-\beta'\mathbf{x}=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}\{[S(5)\mathbf{a}+S(3)\mathbf{b}+S(2)c]\mathbf{x}+[S(4)\mathbf{a}+S(2)\mathbf{b}+S(1)\mathbf{c}]\mathbf{x}_D'\},$$
$$\mathbf{y}_0'-\beta'\mathbf{y}=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}\{[S(5)\mathbf{a}+S(3)\mathbf{b}+S(2)\mathbf{c}]\mathbf{y}+[S(4)\mathbf{a}+S(2)\mathbf{b}+S(1)\mathbf{c}]\mathbf{y}_D'\}. \quad (4,2)$$

We can now discuss the individual coefficients in the usual way.

We consider first the rays through the middle of the object, defined by the values: $\mathbf{a}=\mathbf{b}=0=\mathbf{x}=\mathbf{y}$.

We find

$$\mathbf{x}_0'=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}S(1)\mathbf{c}\mathbf{x}_D',$$
$$\mathbf{y}_0'=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}S(1)\mathbf{c}\mathbf{y}_D'. \quad (4,3)$$

If $S(1)=0$, all rays from the center of the object would join at the Gaussian image point. Otherwise rays intersect the plane through the Gaussian point in a circle whose radius is proportional to $S(1)$. $S(1)$ is called the error of *spherical aberration*. A better designation for this would be *aperture error*, as this aberration of course occurs also in aspherical systems.

For the rays through the center of the exit pupil (so-called principal rays), which are defined by: $\mathbf{x}'_D=\mathbf{y}_D'=0=\mathbf{b}=\mathbf{c}$, we see that

$$\mathbf{x}_0'=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}S(5)\mathbf{a}\mathbf{x},$$
$$\mathbf{y}_0'=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}S(5)\mathbf{a}\mathbf{y}. \quad (4,4)$$

According to Gullstrand, the principal rays give an optical projection of the object plane on the image plane. This optical projection would be similar if $S(5)=0$. If $S(5)\neq 0$, there will no longer be a similar point-to-point correspondence, but, on the contrary, the magnification will be a function of the distance of the object point from the axis, and will increase or decrease according to the sign of $S(5)$. This is known as the *error of distortion*.

We now shall consider an object which is sufficiently small to permit us to neglect all terms of (4, 2) which have $\mathbf{x}$, $\mathbf{y}$ in a power greater than the first. We have left then only two terms:

$$\mathbf{x}_0'-\beta'\mathbf{x}=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}\{S(2)\mathbf{c}\mathbf{x}+S(2)\mathbf{b}\mathbf{x}_D'+S(1)\mathbf{c}\mathbf{x}_D'\},$$
$$\mathbf{y}_0'-\beta'\mathbf{y}=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}\{S(2)\mathbf{c}\mathbf{y}+S(2)\mathbf{b}\mathbf{y}_D'+S(1)\mathbf{c}\mathbf{y}_D'\}, \quad (4,5)$$

namely, $S(1)$, the aperture aberration, and $S(2)$, which is called the *error of coma*. If $S(2)=0$, to every point of the object would correspond a circle of confusion of the same size as the circle of confusion for the axial point. In this case, all neighboring caustics would be rotationally symmetrical, as is the caustic belonging to the axial point of the object, and the object would have a uniform image. Coma destroys this uniformity. It is easy to investigate the effects of this error alone: merely let $S(1)=0$ in (4, 5). In this manner we verify the well-known effects of coma.

We have remaining now the task of discussing the *image field errors*, $S(3)$ and $S(4)$. This is done most easily in the following manner: We make the diaphragm so small that we may neglect $\mathbf{x}_D'$, $\mathbf{x}_D'$ in powers greater than one, and consider the rays coming from a fixed object point $\mathbf{x}$, $\mathbf{y}$, off the axis. Eq. (4, 4) gives for the point at which a ray intersects the plane through the Gaussian image point:

$$\mathbf{x}_0'-\beta'\mathbf{x}-\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}S(5)\mathbf{a}\mathbf{x}=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}\times\{S(3)\mathbf{b}\mathbf{x}+S(4)\mathbf{a}\mathbf{x}_D'\},$$
$$\mathbf{y}_0'-\beta'\mathbf{y}-\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}S(5)\mathbf{a}\mathbf{y}=\frac{1}{2}\frac{\beta'\mathbf{C}_D}{\beta_D'D_1}\times\{S(3)\mathbf{b}\mathbf{y}+S(4)\mathbf{a}\mathbf{y}_D'\}. \quad (4,6)$$

After an easy development, (6) shows that, in general, the rays from an extra-axial object point passing through a small diaphragm opening meet the image plane in an ellipse, whose axes are proportional to

$$S(4) : 2S(3) + S(4) \quad (4,7)$$

respectively. The center of the ellipse is the point where the principal ray intersects the image plane.

This error is due to the fact that we have, in general, not corrected the astigmatism along the principal rays. Therefore, to each object point corresponds a meridional image point, formed by the rays in the meridian plane, and a sagittal image point, formed by the rays in the orthogonal plane.

This may also be described in the following way: The rays, coming from an extra-axial object point and leaving the system, are, in general, a set of normals to an arbitrary set of wave surfaces, which are of double curvature. The centers of curvature are the meridional and sagittal image points. It is obvious that the entity of rays surrounding the principal ray must meet the image plane in an ellipse whose axes are proportional to the distances of the two image points from the plane.

We call

$$\begin{aligned} &S(4) \text{ the sagittal curvature,} \\ &2S(3)+S(4) \text{ the meridional curvature.} \end{aligned} \tag{4, 8}$$

Now we see that if $S(3)=0$, both curvatures are equal; in which case the rays intersect the image plane in a circle, and the sagittal and meridional focal points fall together. Then we say the system is free from astigmatism. For this reason, we call

$$S(3) \textit{ the astigmatic error.} \tag{4, 9}$$

We can see a little more about these errors if, according to a suggestion made by the author,[6] we develop and investigate them for a curved object and a curved surface through the Gaussian image point.

Let us designate the coordinates of the object point by $\mathbf{x}^*$, $\mathbf{y}^*$, $\mathbf{z}^*$, analog $\mathbf{a}^*$, $\mathbf{b}^*$, and the coordinates of the image point by $\mathbf{x}_0'^*$, $\mathbf{y}_0'^*$, $\mathbf{z}_0'^*$. Then we get easily equations corresponding to (4, 2) for the new coordinates.

In these equations, the values of the new errors depend on the errors for plane object and image, and the curvatures ($\mathbf{R}^*$, $\mathbf{R}'^*$) of the object and image surfaces. They are related:

$$\begin{aligned} &S_1^*=S_1, \quad S_4^*=S_4+\mathbf{R}'^*-\mathbf{R}^*, \\ &S_2^*=S_2, \quad S_5^*=S_5+\beta'\mathbf{R}'^*-\beta_D'\mathbf{R}^*. \\ &S_3^*=S_3, \end{aligned} \tag{4, 10}$$

We need not define the image errors, for they have the same sense for the curved object and image, except that the meaning of distortion is slightly changed. If $S^*(5)=0$, according to (4, 4), $\mathbf{x}'^*$, $\mathbf{y}'^*$ and $\mathbf{x}^*$, $\mathbf{y}^*$ are proportional. We suggest saying that, if $S^*(5)=0$, the object has a "*distance-true*" image, instead of speaking of an undistorted image.

Let us now discuss Eq. (4, 10). For curved object and image, we see that the image field errors give us, for every point off the axis, an ellipse, whose principal axes are proportional to

$$\begin{aligned} &S_4^*=S_4+\mathbf{R}'^*-\mathbf{R}^*, \\ &2S_3^*+S_4^*=2S_3+S_4+\mathbf{R}'^*-\mathbf{R}^*. \end{aligned} \tag{4, 11}$$

We investigate first the case when $S_3=0$, which means that the astigmatism vanishes. Then we see that, for each object surface (curvature $\mathbf{R}^*$), we can determine an image surface (curvature $\mathbf{R}'^*$), such that $S_4^*=2S_3^*+S_4^*=0$. This surface is the anastigmatic image of $\mathbf{R}^*$ for the principal rays. Between R^* and R'^* there is this relation:

$$\mathbf{R}'^*=\mathbf{R}^*-S_4. \tag{4, 12}$$

The intersections with all other surfaces through the Gaussian image point are circles.

For every surface through the object point there is a corresponding unique anastigmatic image surface. But only a certain pair of these surfaces has a distance-true image. For, in that case, S_5^* must also vanish. The curvatures of object and image surfaces for this distance-true pair are given by

$$\begin{aligned} &\mathbf{R}^*=(\beta'S_4-S_5)/(\beta'-\beta_D'), \\ &\mathbf{R}'^*=(\beta_D'S_4-S_5)/(\beta'-\beta_D'). \end{aligned} \tag{4, 13}$$

The author supposes that, if a finite field has an anastigmatic image, the image is always distance-true in the neighborhood of the axis. This means that the two surfaces of (13) are distinguished from all others because they lack astigmatism.

If $S_3 \neq 0$, we recognize that we can find for

[6] M. Herzberger, *Strahlenoptik* (Verlag J. Springer, Berlin, 1931).

every object surface, two image surfaces, which are such that the ellipse given by (11) degenerates into a line.

If we choose

$$\mathbf{R}'^{*}=\mathbf{R}^{*}-S_4 \qquad (4, 14)$$

and

$$\mathbf{R}'^{*}=\mathbf{R}^{*}-(2S_3+S_4) \qquad (4, 15)$$

respectively, we have these surfaces, in each of which there is a line element corresponding to each object point; in the first, a line element situated in the meridional plane, the other situated in the orthogonal plane. In this case, we usually say that an object surface has two corresponding image surfaces: the sagittal surface and the meridional surface.

According to (14), $\mathbf{R}'^{*}$ is called the sagittal curvature, and according to (15), the meridional curvature corresponding to $\mathbf{R}^{*}$.

The figure of confusion is a circle if

$$\mathbf{R}'^{*}=\mathbf{R}^{*}-(S_3+S_4), \qquad (4, 16)$$

for in this case we have:

$$S_4^{*}=-(2S_3^{*}+S_4^{*}).$$

The curvature given by (16) is known as the average curvature of the image field.

Now Seidel has given formulae for the five image errors. He used the pole-to-pole method of calculation, and in this system, the magnitude:

$$Q_\nu=n/s-n/r=n'/s'-n'/r \qquad (4, 17)$$

is invariant.

The standardization of the Seidel coefficients by means of (4, 2) now leads to Seidel's well-known formulae:

$$S(1)=\frac{s_1^{4}}{n^2\beta_D'^{2}(s_1-s_{1,D})^2}\sum\left(\frac{h_\nu}{h_1}\right)^4 Q_\nu^{2}\Delta\mathbf{S},$$

$$S(2)=\frac{s_1^{3}s_{1,D}}{n^2\beta_D'(s_1-s_{1,D})^2}\sum\left(\frac{h_\nu}{h_1}\right)^3\left(\frac{h_\nu}{h_1}\right)_D Q_\nu Q_{\nu,D}\Delta\mathbf{S},$$

$$S(3)=\frac{s_1^{2}s_{1,D}^{2}}{n^2(s_1-s_{1,D})^2}\sum\left(\frac{h_\nu}{h_1}\right)^2\left(\frac{h_\nu}{h_1}\right)_D^2 Q_{\nu,D}^{2}\Delta\mathbf{S}, \qquad (4, 18)$$

$$S(4)=\frac{s_1^{2}s_{1,D}^{2}}{n^2(s_1-s_{1,D})^2}\sum\left(\frac{h_\nu}{h_1}\right)^2\left(\frac{h_\nu}{h_1}\right)_D^2 \times Q_{\nu,D}^{2}\Delta\mathbf{S}-\sum\Delta\mathbf{R},$$

$$S(5)=\frac{s_1s^{2}{}_{1,D}\beta_D'}{n^2(s_1-s_{1,D})^2}\sum\left(\frac{h_\nu}{h_1}\right)\left(\frac{h_\nu}{h_1}\right)_D^3\frac{Q_{\nu,D}^{3}}{Q_\nu}\Delta\mathbf{S}.$$

$$-\frac{s_{1,D}\beta_D'}{s_1}\sum\frac{(h_\nu/h_1)_D}{(h_\nu/h_1)}\frac{Q_{\nu,D}}{Q_\nu}\Delta\mathbf{R}.$$

We note

$$\frac{h_\nu}{\mathbf{p}_\nu}=\frac{s_\nu\sigma_\nu}{\mathbf{c}_\nu\sigma_\nu}=\frac{(\mathbf{c}_\nu+\mathbf{r}_\nu)r_\nu}{\mathbf{c}_\nu\mathbf{r}_\nu}=\frac{\mathbf{r}_\nu}{n_\nu}I_\nu=r_\nu I_\nu, \qquad (4, 19)$$

$$h_\nu=\mathbf{r}_\nu I_\nu\mathbf{p}_\nu\cdot(1/n_\nu),$$

$$Q_\nu=n/s-n/r=-\mathbf{c}/r(c+r)=-n_\nu^{2}/\mathbf{r}_\nu^{2}I_\nu, \qquad (4, 20)$$

$$\Delta\mathbf{S}=\Delta(1/\mathbf{c}+\mathbf{r})=(1/I)\Delta\mathbf{CR}. \qquad (4, 21)$$

If we insert this in (18) we obtain

$$S_1=\left(\frac{\mathbf{C}_{1,D}}{\beta_D'\mathbf{C}_\nu'}\right)^2\frac{1}{D_1^{2}}\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)^4 I_\nu\Delta_\nu\mathbf{RC},$$

$$S_2=\frac{\mathbf{C}_{1,D}}{\beta_D'\mathbf{C}_\nu'}\cdot\frac{1}{D_1^{2}}\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)^3\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D I_\nu\Delta_\nu\mathbf{RC},$$

$$S_3=\frac{1}{D_1^{2}}\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)^2\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^2 I_\nu\Delta_\nu\mathbf{RC}, \qquad (4, 22)$$

$$S_4=\frac{1}{D_1^{2}}\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)^2\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^2[I_\nu\Delta_\nu\mathbf{RC}-\Delta_\nu\mathbf{R}],$$

$$S_5=\frac{\beta_D'\mathbf{C}_\nu'}{\mathbf{C}_{1,D}}\cdot\frac{1}{D_1^{2}}\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^3[I_\nu\Delta_\nu\mathbf{RC}-\Delta_\nu\mathbf{R}].$$

We note that these formulae are identical with (18) but easier to calculate. We add, for reasons which will better be understood in the next section, the aperture (spherical) aberration of the stop, as a sixth error.

The corresponding formula is obviously

$$S_6=\left(\frac{\beta_D'\mathbf{C}_\nu'}{\mathbf{C}_{1,D}}\right)^2\cdot\frac{1}{D_1^{2}}\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^4 I_D\Delta(\mathbf{RC}_D). \qquad (4, 23)$$

Section V. Development of the Seidel Aberration Errors in Reference to the Position of Stop and Object

Seidel has demonstrated a method of developing the image errors in reference to the position of the stop. We apply this method to our formulae (4, 22). We must remark, however, that

the partial coefficients obtained in this manner are not the same as those which Seidel developed. They are simpler, and, in the author's opinion, better suited to give the true contribution of each surface to the image error. If we insert

$$\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D = \frac{\mathbf{p}_\nu}{\mathbf{p}_1}(1+D_1 Z_\nu) \tag{5, 1}$$

in (4, 22), we get

$$S(1) = \left(\frac{\mathbf{C}_{1,D}}{\beta_D'\mathbf{C}_\nu'}\right)^2 \frac{1}{D_1^2}(A),$$

$$S(2) = \frac{\mathbf{C}_{1,D}}{\beta_D'\mathbf{C}_\nu'}\frac{1}{D_1^2}[(A)+D_1(B)],$$

$$S(3) = \frac{1}{D_1^2}[(A)+2D_1(B)+D_1^2(C)], \tag{5, 2}$$

$$S(4) = \frac{1}{D_1^2}[(A)+2D_1(B)+D_1^2(D)],$$

$$S(5) = \frac{\beta_D'\mathbf{C}_\nu'}{\mathbf{C}_{1,D}}\frac{1}{D_1^2}[(A)+3D_1(B) + D_1^2\{2(C)+(D)\}+D_1^3(E)],$$

where

$$(A) = \sum(A)_\nu = \sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)^4 I_\nu \Delta\mathbf{RC},$$

$$(B) = \sum(B)_\nu = \sum(A)_\nu Z_\nu,$$

$$(C) = \sum(C)_\nu = \sum(B)_\nu Z_\nu = \sum(A)_\nu Z_\nu^2, \tag{5, 3}$$

$$(D) = \sum(D)_\nu = \sum(A)_\nu Z_\nu^2 - \sum\Delta\mathbf{R},$$

$$(E) = \sum(E)_\nu = \sum(D)_\nu Z_\nu = \sum(A)_\nu Z_\nu^3 - \sum Z_\nu \Delta\mathbf{R}$$

and

$$Z_\nu = \sum \frac{\mathbf{m}_{\nu,\,\nu-1}}{(\mathbf{p}_{\nu-1}/\mathbf{p}_1)(\mathbf{p}_\nu/\mathbf{p}_1)}. \tag{5, 4}$$

From (2) we see that the aperture aberration, S_1, vanishes, if $(A)=0$. Then we say the object point is spherically corrected.*

In this case, then $S(2)$ can vanish only if $(B)=0$, and then $S(2)$ disappears independently of the stop position. A point where $S(2)=S(1)=0$ is called an *aplanatic* object point. Every surface element through an aplanatic point has an image free of coma.

If the object point is aplanatic, the astigmatism can be corrected only if $(C)=0$. Then, however, the astigmatism is corrected for any position of the stop. Such an image we call a *near-field-sharp* image.

We know from (4, 11) that in this case every surface through the object point has a sharp image, but only the one given by (4, 13) has an image which is also distance-true.

In a *near-field-sharp* point, the error $S(4)$ cannot be corrected unless $(D)=0$, but then the correction is independent of the position of the stop. In this case, the flat object through the origin has a flat image. We call such an image a *near-flat-sharp* image.

In a *near-flat-sharp* image, $S(5)$ can be corrected only if $(E)=0$, but it will then be corrected for all positions of the stop. This we speak of as a *perfect image.*

In all other cases, there is one position of the stop to correct coma, two positions to correct astigmatism or curvature, and three positions which correct distortion.

Interesting from the optical standpoint is the case when the stop position for astigmatic correction, and for coma correction are at the same place. Then we call the image *near-field-equal.* The condition is, according to (2)

$$(A)(C) = (B)^2. \tag{5, 5}$$

Zincken-Sommer, a step-son of Voigtländer, first paid attention to the importance of this condition. If we can regard the condition of aplanatism as the Seidel counterpart of the well-known sine condition, we have here the counterpart of the Staeble-Lihotzky isoplanatic condition for *near-field-equal* image. Eq. (2) shows that in this case there exists only one anastigmatic stop.

Finally, we shall develop $S(6)$ according to the stop. Eq. (3, 2) gives

$$I_D\Delta\mathbf{C}_D\mathbf{R} = I\Delta\mathbf{CR} + D\Delta\mathbf{C}^2 + D^2\Delta\mathbf{R}. \tag{5, 6}$$

If now we remember 2, (6) and (2, 10), we get

$$S(6) = \left(\frac{\beta_D'\mathbf{C}_\nu'}{\mathbf{C}_{1,D}}\right)^2 \frac{1}{D_1^2}\{(1+D_1Z_\nu)^4\sum I_\nu\Delta\mathbf{CR} + D_1(1+D_1Z_\nu)^3\Delta\mathbf{C}^2 + D_1^2(1+D_1Z_\nu)^2\Delta\mathbf{R}\} \tag{5, 7}$$

* All definitions are given with regard to the Seidel theory.

or, according to (3, 7)

$$S(6)=\left(\frac{\beta_D'\mathbf{C}_\nu'}{\mathbf{C}_{1,\,D}}\right)^2\frac{1}{D_1^2}\{(A)+D_1[4(B)+\rho_0]+D_1^2[2(2(C)+(D))+3\sigma_0]+D_1^3[4(E)+3\tau_0]+D_1^4(F)\} \quad (5,\,8)$$

where

$$(F)=\sum Z_\nu(E)_\nu+\sum Z_\nu^2(Z_\nu\Delta\mathbf{C}^2+2\Delta\mathbf{R}). \quad (5,\,9)$$

We shall now make the same development with reference to the object. We rewrite, from (3, 2), the two equations:

$$\begin{aligned} I\Delta\mathbf{CR}&=I_D\Delta\mathbf{C}_D\mathbf{R}-D\Delta\mathbf{C}_D^2+D^2\Delta\mathbf{R},\\ I\Delta\mathbf{CR}-D^2\Delta\mathbf{R}&=I_D\Delta\mathbf{C}_D\mathbf{R}-D\Delta\mathbf{C}_D^2, \end{aligned} \quad (5,\,10)$$

and, remembering (3, 7)

$$\begin{aligned} \rho_D&=\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^2 Z_{\nu,\,D}^2\Delta\mathbf{C}_D^2-2Z_{\nu,\,D}\Delta\mathbf{R},\\ \sigma_D&=\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^2 Z_{\nu,\,D}\Delta\mathbf{C}_D^2-\Delta\mathbf{R},\\ \tau_D&=\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^2\Delta\mathbf{C}_D^2. \end{aligned} \quad (5,\,11)$$

Thus, if we introduce analogously to (5, 3) these easily understood abbreviations:

$$\begin{aligned} (A)_D&=\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^4 I_D\Delta\mathbf{C}_D\mathbf{R},\\ (B)_D&=\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^4 Z_\nu I_D\Delta\mathbf{C}_D\mathbf{R},\\ (C)_D&=\sum\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D^4 Z_\nu^2 I_D\Delta\mathbf{C}_D\mathbf{R}, \end{aligned} \quad (5,\,12)$$

$$\begin{aligned} (D)_D&=(C)_D-\sum\Delta\mathbf{R}=\sum(D)_{\nu,\,D},\\ (E)_D&=\sum(D)_{\nu,\,D}Z_{\nu,\,D}=\sum(E)_{\nu,\,D},\\ (F)_D&=\sum(E)_{\nu,\,D}Z_{\nu,\,D}\\ &\quad+\sum Z_{\nu,\,D}^2(Z_{\nu,\,D}\Delta\mathbf{C}_D^2-2\Delta\mathbf{R}), \end{aligned}$$

into (4, 22) we get

$$\begin{aligned} S(1)&=\left(\frac{\mathbf{C}_{1,\,D}}{\mathbf{C}_\nu'\beta_D'}\right)^2\frac{1}{D_1^2}\{(A)_D-D_1[4(B)_D+\tau_D]\\ &\quad+D_1^2[2(2(C)_D+(D)_D)+3\sigma_D]\\ &\quad-D_1^3[4(E)_D+3\rho_D]+D_1^4(F)_D\},\\ S(2)&=\left(\frac{\mathbf{C}_{1,\,D}}{\mathbf{C}_\nu'\beta_D'}\right)\frac{1}{D_1^2}\{(A)_D-D_1[2(B)_D+\tau_D]\\ &\quad+D_1^2[2(C)_D+(D)_D+2\sigma_D]\\ &\quad-D_1^3[(E)_D+\rho_D]\},\\ S(3)&=\frac{1}{D_1^2}\{(A)_D-D_1[2(B)_D+\tau_D]\\ &\quad+D_1^2[(C)_D+\sigma_D]\},\\ S(4)&=\frac{1}{D_1^2}\{(A)_D-D_1[2(B)_D+\tau_D]\\ &\quad+D_1^2[(D)_D+\sigma_D]\},\\ S(5)&=\frac{\mathbf{C}_\nu'\beta_D'}{\mathbf{C}_{1,\,D}}\cdot\frac{1}{D_1^2}\{(A)_D-D_1[(B)_D+\tau_D]\},\\ S(6)&=\left(\frac{\mathbf{C}_\nu'\beta_\nu'}{\mathbf{C}_{1,\,D}}\right)^2\frac{1}{D_1^2}(A)_D. \end{aligned} \quad (5,\,13)$$

Comparison of (5, 13) with (5, 2) shows, if we remember (3, 6)

$$\begin{aligned} \rho&=\tau_D-2D_1\sigma_D+D_1^2\rho_D,\\ \sigma&=\tau_D-D_1\sigma_D,\quad \tau=\tau_D. \end{aligned} \quad (5,\,14)$$

We put for abbreviation

$$K=\mathbf{C}_{1,\,D}/\mathbf{C}_\nu'\beta_D' \quad (5,\,15)$$

and get

$$\begin{aligned} S(1)&=K^4S_D(6),\\ S(2)&=K^2S_D(5)-\frac{K}{D_1}\cdot\rho,\\ S(3)&=S_D(3)-\frac{1}{D_1}\cdot\sigma,\\ S(4)&=S_D(4)-\frac{1}{D_1}\cdot\sigma,\\ S(5)&=\frac{1}{K^2}S_D(2)-\frac{1}{KD_1}\cdot\tau,\\ S(6)&=\frac{1}{K^4}S_D(1). \end{aligned} \quad (5,\,16)$$

or, corresponding

$$\begin{aligned} S_D(1) &= K^4 S(6), \\ S_D(2) &= K^2 S(5) + \frac{K}{D_1}\cdot\tau, \\ S_D(3) &= S(3) + \frac{1}{D_1}\cdot\sigma, \\ S_D(4) &= S(4) + \frac{1}{D_1}\cdot\sigma, \\ S_D(5) &= \frac{1}{K^2}S(2) + \frac{1}{D_1 K}\cdot\rho, \\ S_D(6) &= \frac{1}{K^4}S(1). \end{aligned} \qquad (5, 16a)$$

Eq. (13) shows us: In a general system, spherical aberration can be corrected for four points only; there are, in general, three object points for which a given stop gives a coma-free image, two object points for which it gives an image free from astigmatism, and only a single object for which the given stop gives an image free from distortion.

Eqs. (13)–(15) show us how to investigate systems with special qualities. That will be the task of the following section.

Here we call the attention of the reader only to the fact that

$$S(4) - S(3) = S_D(4) - S_D(3) = -\sum \Delta \mathbf{R} \qquad (5, 17)$$

is an invariant for the system. Eq. (17) is called the Petzval condition, after Petzval, who was the first to demonstrate its importance.

If we have a curved object and image, Eq. (4, 10) shows that this is valid:

$$S^*(4) - S^*(3) = \mathbf{R}'^* - \mathbf{R}^* - \sum \Delta \mathbf{R}. \qquad (5, 18)$$

We see that, if there is an anastigmatic image, we must have

$$S^*(4) = S^*(3) = 0,$$

which means

$$\mathbf{R}'^* = \mathbf{R}^* + \sum \Delta \mathbf{R}. \qquad (5, 19)$$

Eq. (19) shows us the condition that must be fulfilled between the curvature of object and image, if the image is to be anastigmatic. Eq. (19) is a necessary condition, but not a sufficient one. The object and image can be planes only when the Petzval sum vanishes.

Section VI. Special Systems

We shall investigate first systems for which every point is without spherical aberration. Sir W. Herschel suggested that telescopic systems be corrected so that every point of the axis is spherically corrected. Therefore, we call such systems

(a) Herschel systems

A Herschel system, according to (5, 13), is given by:

$$\begin{aligned} (A)_D &= 0, & (E)_D &= -3\rho_D/4, \\ (B)_D &= -\tau_D/4, & (F)_D &= 0. \\ 2(C)_D + (D)_D &= -3\sigma_D/2, \end{aligned} \qquad (6, 1)$$

(1) is a necessary and sufficient condition that a system is a Herschel system. We shall investigate the properties of a Herschel system. We put (1) into (5, 13), and have

$$S(1) = 0,$$

$$S(2) = -\frac{K}{D_1}\left[\frac{\tau_D}{4} - \frac{\sigma_D}{2}D_1 + \frac{\rho_D}{4}D_1^2\right] = -\frac{K}{4D_1}\cdot\rho,$$

$$2S(3) + S(4) = -\frac{1}{D_1}\left[\frac{3\tau_D}{2} - \frac{3\sigma_D}{2}D_1\right] = -\frac{3}{2D_1}\sigma, \qquad (6,2)$$

$$S(5) = -\frac{1}{KD_1}\cdot\frac{3\tau_D}{4} = -\frac{3}{4KD_1}\cdot\tau.$$

In the Herschel system, the nodal points are the only aplanatic points. The nodal points also determine the only positions for orthoscopic diaphragms. Orthoscopic means that all objects are free from distortion. If the Herschel system is not a nodal system (a telescopic system with $\beta' = 1$), that is, if

$$\rho = \sigma = \tau = 0 \qquad (6, 3)$$

does not hold, then there is no object with corrected tangential astigmatism, and, therefore, no object with a perfect image anywhere.

If (3) is fulfilled, every object has a perfect image.

(b) The triaplanatic point systems

From (6, 2) we learned that in a system which is not a nodal system, it is not possible for all points to be aplanatic. Eq. (5, 13) shows that a general system cannot have more than three aplanatic points. Systems which have three aplanatic points are called triaplanatic-point systems. We shall investigate the properties of such systems.

If there are three aplanatic points in our system, the two equations:

$$S(1)=0, \quad S(2)=0,$$

or:

$$\frac{1}{K}S(1)-S(2)=0, \quad S(2)=0,$$

must have the same solutions. That means that the coefficients of

$$\begin{aligned}&(B)_D-D_1[2(C)_D+(D)_D+\sigma_D]\\&\quad+D_1^2[3(E)_D+2\rho_D]-D_1^3(F)_D=0,\\&(A)_D-D_1[3(B)_D+\tau_D]+D_1^2[2(C)_D\\&\quad+(D)_D+2\sigma_D]-D_1^3[(E)_D+\rho_D]=0\end{aligned} \tag{6, 4}$$

must be proportional.

The existence of a number λ, such that

$$\begin{aligned}(A)_D&=\lambda(B)_D,\\3(B)_D+\tau_D&=\lambda[2(C)_D+(D)_D+\sigma_D],\\2(C)_D+(D)_D+2\sigma_D&=\lambda[3(E)_D+2\rho_D],\\(E)_D+\rho_D&=\lambda(F)_D,\end{aligned} \tag{6, 5}$$

are simultaneously valid, is a necessary and sufficient condition that the system is triaplanatic.

Eq. (5) holds good independently of where we have chosen the stop. We shall now show how to choose it so that the equations simplify.

The spherical aberration, according to (13), vanishes in four points, of which only three are aplanatic. Now, we can put our origin at the fourth point, where the spherical aberration is corrected, but not the sine condition. At that point, $(A)_D=0$, but $(B)_D\neq 0$, and, because (5) must be valid, notwithstanding the position of the stop, that means that $\lambda=0$.

Now, if we put the stop at that point, we get

$$\begin{aligned}(A)_D&=0, & 2(C)_D+(D)_D&=-2\sigma_D\\(B)_D&=-\tau_D/3, & (E)_D&=-\rho_D.\end{aligned} \tag{6, 6}$$

And putting these values into (5, 13) gives

$$\begin{aligned}S(1)&=\frac{K^2}{D_1}\left[\frac{\tau_D}{3}-\sigma_D D_1+\rho_D D_1^2-D_1^3(F)_D\right]\\&=\left[\rho-\sigma+\frac{\tau}{3}-D_1^3(F)_D\right],\end{aligned}$$

$$S(2)=0, \tag{6, 7}$$

$$2S(3)+S(4)=-\frac{1}{D_1}(\tau_D-D_1\sigma_D)=-\frac{1}{D_1}\sigma,$$

$$S(5)=-\frac{1}{KD_1}\cdot\frac{\tau_D}{3}=-\frac{1}{KD_1}\cdot\frac{\tau}{3}.$$

These equations show us:

In a triaplanatic system there is a stop which gives an image free of coma for all object points. We call such a diaphragm a *spatial coma-free* diaphragm.

Of course, we can also invert this theorem. If a system has a *spatial coma-free* diaphragm, there are three aplanatic object points, which are found by solving the cubic equation $S(1)=0$, according to (6, 7). Of course, all three of these aplanatic points need not be real.

None of the aplanatic points can have the meridional field error corrected, unless the spatial coma-free stop and one of the aplanatic points are reciprocal points. None of the aplanatic points can have a perfect image (unless the system is a nodal system). The spatial coma-free stop is orthoscopic only if it is situated at a nodal point.

We investigate this special case of the triaplanatic system by setting

$$\tau=0. \tag{6, 8}$$

We get

$$\begin{aligned}&S(1)=-K^2[\sigma_D-\rho_D D_1+(F)_D D_1^2],\\&S(2)=0, \qquad S(4)=(D)_D+\sigma_D,\\&S(3)=(C)_D+\sigma_D, \quad S(5)=S(6)=0,\end{aligned} \tag{6, 9}$$

and

$$2S(3)+S(4)=2(C)_D+(D)_D+3\sigma_D=\sigma.$$

In this case, the *spatial coma-free* diaphragm is at the same time the orthoscopic stop. The spatial coma-free stop is itself an aplanatic point, and a double point for correction of spherical aberration. All astigmatic errors are the same for every object, especially if

$$\begin{aligned} &(C)_D=-\sigma_D, \quad (D)_D=0 \\ \text{or}\quad &(D)_D=-\sigma_D, \quad (C)_D=-\sigma_D/2 \\ \text{or}\quad &[(C)_D+(D)_D]/2=-\sigma_D, \\ &(C)_D=0, \quad (D)_D=-2\sigma_D. \end{aligned} \tag{6, 10}$$

Then our two aplanatic points are at once either *near-field-sharp*, or sagittally corrected, or the average curvature is corrected; and the stop is either spatial anastigmatic, or free of sagittal curvature, or gives the average curvature zero for all objects. For instance, this first case is fulfilled at a single surface, where the center is at the same time the spatial coma-free and the spatial anastigmatic diaphragm.

Such a system, according to (6) and (10), is given by

$$\begin{aligned} &(A)_D=(B)_D=(D)_D=0, \quad (C)_D=-\sigma_D, \\ &(E)_D=-\rho_D, \quad \tau_D=0 \end{aligned} \tag{6, 11}$$

$$\begin{aligned} \text{or}\quad &S(1)=-K^2[\sigma_D-\rho_D D_1+(F)_D D_1^2], \\ &S(2)=0, \quad S(3)=0, \; S(4)=\sigma_D=\sigma, \\ &S(5)=0, \quad S(6)=0. \end{aligned} \tag{6, 12}$$

(c) The duo-near-field-sharp-point systems

If our system has a near-field-sharp object point, then the three equations:

$$S(1)=0, \quad S(2)=0, \quad S(3)=0 \tag{6, 13}$$

or, which is the same, these three equations

$$\begin{aligned} &(A)_D-D_1[2(B)_D+\tau_D]+D_1^2[(C)_D+\sigma_D]=0, \\ &(B)_D-D_1[(C)_D+(D)_D+\sigma_D] \\ &\qquad +D_1^2[(E)_D+\rho_D]=0, \\ &(C)_D-D_1[2(E)_D+\rho_D]+D_1^2(F)_D=0 \end{aligned} \tag{6, 14}$$

must have a common solution. Eq. (14) shows that a system (excepting only the nodal system) can have at most two near-field-sharp object points.

We shall now investigate systems with two near-field-sharp object points. In this case, all three of Eqs. (14) must be proportional. That is, between the coefficients A–E, these four equations hold

$$\begin{aligned} &\frac{(A)_D}{(B)_D}=\frac{2(B)_D+\tau_D}{(C_D)+(D)_D+\sigma_D}=\frac{(C)_D+\sigma_D}{(E)_D+\rho_D}, \\ &\frac{(B)_D}{(C)_D}=\frac{(C)_D+(D)_D+\sigma_D}{2(E)_D+\rho_D}=\frac{(E)_D+\rho_D}{(F)_D}. \end{aligned} \tag{6, 15}$$

Eq. (6, 15) is a necessary and sufficient condition that the optical system is a duo-near-field-sharp-point system. If (15) is fulfilled, we get the position of our near-field-sharp objects by solving one of Eqs. (14) for D_1.

In order to investigate the properties of the duo-near-field-sharp-point system, we shall now put our object and stop at the two near-field-sharp points. We have then

$$\begin{aligned} &S(1)=S(2)=S(3)=0, \\ &S_D(1)=S_D(2)=S_D(3)=0. \end{aligned} \tag{6, 16}$$

Now, putting (16) into (5, 16), we find

$$\begin{aligned} &S_D(5)=\frac{1}{KD_1}\cdot\rho, \quad S(5)=-\frac{1}{KD_1}\cdot\tau, \\ &\sigma=0, \qquad\qquad S(4)=S_D(4). \end{aligned} \tag{6, 17}$$

Eq. (17) shows first that $\sigma=0$. That means:

If an optical system has two near-field-sharp object points, these two points must be situated in reciprocal points; that is, they must have reciprocal reduced magnification.

We can use the other equations to determine the objects which are distance-true images of each other. We find, according to (4, 13)

$$\begin{aligned} &\mathbf{R}_1^*=\frac{1}{\beta'-1/\beta'}\left[\beta' S(4)+\frac{1}{KD_1}\cdot\tau\right], \\ &\mathbf{R}_1'^*=\frac{1}{\beta'-1/\beta'}\left[\frac{1}{\beta'}S(4)+\frac{1}{KD_1}\cdot\tau\right], \\ &\mathbf{R}_2^*=-\frac{1}{\beta'-1/\beta'}\left[\frac{1}{\beta'}S(4)-\frac{1}{KD_1}\cdot\rho\right], \\ &\mathbf{R}_2'^*=-\frac{1}{\beta'-1/\beta'}\left[\beta' S(4)-\frac{1}{KD_1}\cdot\rho\right]. \end{aligned} \tag{6, 18}$$

(d) A special system, with astigmatic errors corrected for two points

It might be of special interest to see what happens if a system has two aplanatic points which are corrected for other field errors than astigmatism, for instance, the meridional field curvature

$$2S(3)+S(4)=0 \tag{6, 19}$$

or the sagittal field curvature

$$S(4)=0 \tag{6, 20}$$

or the average curvature

$$S(3)+S(4)=0. \tag{6, 21}$$

In any of these cases, (5, 16) shows that, if this requirement is fulfilled for two points, the equation

$$\sigma=0 \tag{6, 22}$$

is always true. That is, in any case, both object points must have a reciprocal reduced magnification. All other conclusions are similar to those drawn for astigmatism, so we will not repeat the corresponding formulae, which are of less theoretical and practical interest.

(e) The uno-perfect-point system

If we look at Eq. (5, 13), we see that there can be only a single point with a perfect image, if the system is not a trivial nodal system. We must have

$$S(1)=S(2)=S(3)=S(4)=S(5)=0 \tag{6, 23}$$

and we get as a necessary and sufficient condition

$$\begin{aligned}(A)_D-D_1[(B)_D+\tau_D]&=0,\\ (B)_D-D_1[(D)_D+\sigma_D]&=0,\end{aligned} \tag{6, 24}$$

$$(B)_D-D_1[(C)_D+\sigma_D]=0,$$

$$(C)_D-D_1[(E)_D+\rho_D]=0, \quad (E)_D-D_1(F)_D=0,$$

or

$$\frac{(A)_D}{(B)_D+\tau_D}=\frac{(B)_D}{(D)_D+\sigma_D}=\frac{(C)_D}{(E)_D+\rho_D}=\frac{(E)_D}{(F)_D}. \tag{6, 25}$$

$$(C)_D=(D)_D.$$

If we investigate the image formation for another object point, we had best put the diaphragm at the perfect point. In that case, we assume

$$S_D(1)=S_D(2)=S_D(3)=S_D(4)=S_D(5)=0$$

or $$(A)_D=(B)_D=(C)_D=(E)_D=(D)_D=0. \tag{6, 26}$$

If we put this into (5, 13) and (5, 16), we find

$$\begin{aligned}&S(1)=-(K^2/D_1)\{\tau_D-3\sigma_D D_1+3\rho_D D_1^2-(F)_D D_1^3\},\\ &S(2)=-(K/D_1)\rho, \quad S(4)=-(1/D_1)\sigma, \qquad (6, 27)\\ &S(3)=-(1/D_1)\sigma, \quad S(5)=-(1/KD_1)\tau.\end{aligned}$$

We see that, in a uno-perfect-point system, there can be no other point which has a perfect image, or even an anastigmatic image. There can be an aplanatic point only if the nodal point is free of spherical aberration.

Section VII. An Interesting Law for the Image Formation in Optical Systems

The task of the optical designer should now be to investigate all optical systems which give a near-field-sharp image. Among these systems there will be some which give a sharp image not only in the region where the Seidel theory holds, but also for finite aperture and field. This task means solving simultaneously the three equations:

$$S(1)=S(2)=S(3)=0, \tag{7, 1}$$

which are equivalent to

$$\begin{aligned}(A)=\sum(A)_\nu=0, \quad (B)=\sum(A)_\nu Z_\nu=0,\\ (C)=\sum(A)_\nu Z_\nu^2=0.\end{aligned} \tag{7, 2}$$

We now split the summation so that the coefficient for the last surface is on the other side of the equation. We shall designate all summations of the partial system which contains all surfaces but the last, by the sign $\sum'$.

Then we find for the last Kth surface

$$\begin{aligned}(A)_K&=-\sum'(A)_\nu=-(A'),\\ (B)_K&=(A)_K Z_K=-\sum'(B_\nu)\\ &=-\sum'(A)_\nu Z_\nu=-(B'), \qquad (7, 3)\\ (C)_K&=(A)_K Z_K^2=-\sum'(C)_\nu\\ &=-\sum'(A)_\nu Z_\nu^2=-(C').\end{aligned}$$

The image errors of the system without the last surface are denoted by (A'), (B'), (C'). Eq. (7, 3) gives the equation:

$$(A')(C')=(B')^2 \qquad (7,4)$$

or, verbally: if a system is corrected near-field-sharp for an object point, the system without the last surface must be corrected so that the condition of Zincken-Sommer is fulfilled for the rays. That is, that the diaphragm for coma correction and the diaphragm for astigmatic correction shall coincide.

We shall demonstrate further that this diaphragm must be at the center of the last surface. According to (5, 2), we find the coma-free stop for the partial system by choosing D_1 such that

$$\Sigma'(A)+D_1\Sigma'(B)=0. \qquad (7,5)$$

From (5) and (3) we find

$$D_1=-1/Z_K \qquad (7,6)$$

or

$$1+D_1Z_K=0. \qquad (7,7)$$

Comparison with (2, 12) gives

$$\left(\frac{\mathbf{p}_\nu}{\mathbf{p}_1}\right)_D=\frac{\mathbf{p}_\nu}{\mathbf{p}_1}(1+D_1Z_K)=0, \qquad (7,8)$$

which means that the stop is at the center of the last surface, Q.E.D.

We can reverse our theorem: We might have a system and an object for which the Zincken-Sommer condition is fulfilled. If we now add one or several surfaces concentric to the exit pupil of the system, then the whole system gives an image which is, for that particular stop, free from astigmatism and coma. If we have chosen the additional surfaces such that spherical aberration is corrected for the object point, we have at once a near-field-sharp image.

The task of searching for systems which fulfill the three equations of (2) is in this way reduced to finding those which fulfill the single Eq. (4).

The result arrived at in this section seems at first sight very surprising, but on investigation it appears very simple. We shall demonstrate its validity for finite angle and finite aperture.

We assume a system which gives a sharp image of a finite object for a finite aperture. We know that then we can reverse the direction of the ray trace. Therefore, the rays coming from a point on the object surface must form a symmetrical caustic with the axis through the center of the last surface before leaving it. To every point of the object must correspond a symmetrical caustic, and all axes of symmetry meet at the center of curvature of the last surface. This is the simple explanation of our interesting law.

Of course, the first surface must have its center at the entrance pupil for the same reason, because, if we assume a sharp image, object and image may be regarded as interchangeable.

Section VIII. The Concentric Lens System

In the case of a concentric system, some of our formulae simplify.

First, all $\mathbf{m}_\nu$'s are zero, which means

$$Z_\nu=0 \qquad (8,1)$$

for all surfaces.

We have in this case also

$$\mathbf{c}_{\nu+1}=\mathbf{c}_\nu' \qquad (8,2)$$

and therefore, from (1, 17)

$$\mathbf{p}_\nu=\mathbf{p}_{\nu-1}=\cdots\mathbf{p}_1. \qquad (8,3)$$

Thus, we have

$$\begin{aligned}&(A)_\nu=I_\nu\Delta\mathbf{CR}, \quad (D)_\nu=-\Delta\mathbf{R},\\ &(B)_\nu=(C)_\nu=0, \quad (E)_\nu=(F)_\nu=0.\end{aligned} \qquad (8,4)$$

In (4) we may substitute

$$\mathbf{C}_{\nu+1}=\mathbf{C}_\nu'=\mathbf{C}_1+\sum_\nu\Delta\mathbf{R} \qquad (8,5)$$

and we see that A is in this case a *quadratic* function of C.

If we substitute this case under the types already discussed, we see at once that the center is a nodal point, an aplanatic point, and a diaphragm free of coma *and* astigmatism for all object points.

$(A)_\nu=0$, is a quadratic equation in C; if there are two real solutions, there are *two* surfaces which are free of all image errors.

We see, in a concentric system, that if there is one single point free from spherical aberration, there are two surfaces which have sharp images.

The focal length is indicated by $\Phi=-\sum\Delta\mathbf{R}$. If $\Phi=0$, then our system is a nodal system. If

then the spherical aberration is zero for a single point, it is zero for all points, and we have a perfect system which gives a sharp image for all objects, with the reduced magnification of: $\beta' = 1$.

Examples for these theorems are given by Boegehold and Herzberger.[7]

Section IX. The Simplest Optical Systems

(a) A single surface

First we investigate a single optical surface. We have found:

$$\begin{aligned} &(A) = I\,\mathfrak{A}\mathbf{CR} \\ &(B) = 0, \quad (D) = -\mathfrak{A}\mathbf{R}, \\ &(C) = 0, \quad (E) = 0. \end{aligned} \tag{9, 1}$$

We see:

The center of curvature of a single surface is the orthoscopic diaphragm, spatial coma-free, and spatial anastigmatic diaphragm. We have, therefore, two anastigmatic objects, one given by

$$I = 0 \quad \text{or} \quad \mathbf{C} = -\mathbf{R} \tag{9, 2}$$

situated at the pole, and the other given by

$$\mathfrak{A}\mathbf{CR} = 0, \tag{9, 3}$$

which gives us the aplanatic points of the sphere. There is no perfect image, and the calculation of the radii of the object and image which are distance-true leads to the results which have already been found.

The center is obviously the Zincken-Sommer diaphragm for all object points.

(b) Two surfaces

We have, in general

$$\begin{aligned} &(A) = (A)_1 + (A)_2, \quad (B) = (A)_2 Z_2, \\ &(C) = (A)_2 Z_2^2. \end{aligned} \tag{9, 4}$$

[7] H. Boegehold and M. Herzberger, Zeits. f. ang. Math. 15, 157–78 (1935).

We see that our object fulfills the Zincken-Sommer condition only if

$$(A)_1 (A)_2 Z_2^2 = 0. \tag{9, 5}$$

There are three cases:

(1) $A_1 = 0$, i.e., the object point is either at the pole of the first surface, or at the aplanatic point. In this case, the center of curvature of the second surface is the Zincken-Sommer diaphragm for the system.

(2) $A_2 = 0$. The image is at the pole or the aplanatic point of the second surface. The center of the first surface is the Zincken-Sommer diaphragm.

(3) $Z_2 = 0$. The two surfaces of the lens are concentric. In this case the center is the Zincken-Sommer diaphragm for all objects. The system has a spatial coma-free and spatial anastigmatic diaphragm.

(c) Three surfaces

$$\begin{aligned} &(A) = (A)_1 + (A)_2 + (A)_3, \\ &(B) = (A)_2 Z_2 + (A)_3 Z_3, \\ &(C) = (A)_2 Z_2^2 + (A)_3 Z_3^2. \end{aligned} \tag{9, 6}$$

The Zincken-Sommer condition leads to

$$\begin{aligned} (A)_1 [(A)_2 Z_2^2 + (A)_3 Z_3^2] \\ + (A)_2 (A)_3 [Z_3 - Z_2]^2 = 0. \end{aligned} \tag{9, 7}$$

If we want a system with a near-field-sharp object, we get two cases

$$(1) \quad (A)_1 = (A)_2 = (A)_3 = 0 \tag{9, 8}$$

or

$$\begin{aligned} (2) \quad &Z_2 = Z_3, \quad (A)_2 = -(A)_3, \\ &(A)_1 = 0. \end{aligned} \tag{9, 9}$$

In the second case, the two last surfaces are concentric.

Further discussion will be given in a later paper.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

Theory of Microscope Vision*†

M. HERZBERGER, *Eastman Kodak Company*

(Received November 2, 1935)

MICROSCOPIC vision is one of the few places in optics where there are still differences of opinion and vehement disputes. Especially in Germany is there a hard struggle. Since the first theory of microscope vision was given by Ernst Abbe, the founder of the Zeiss Works, and since the most outstanding adversary of the Abbe theory is M. Berek, head of the scientific staff of the Leitz Works, this struggle has sometimes departed from purely scientific discussion. The speaker has had the opportunity of being a member of both staffs, and he has, therefore, been able to make himself acquainted with the principal arguments on both sides.

The speaker will try to show the problems encountered in any theory of microscope vision, and to give a theory which does due justice to both sides. In short, the author believes that the method of Abbe is more far-reaching, and, in one sense, is more correct. The author believes that future progress lies in changing the microscope in such a way that it can better be used as a kind of interferometer, as it does, for instance, in the experiments of the restricted Abbe theory. On the other hand, the author believes that the equivalence principle of M. Berek is a very successful formulation, which enables us to deal very easily with the usual method of using the microscope, and is quite sufficient for this purpose, whereas the Abbe theory is too complicated to supply answers to the questions in daily microscope work.

The speaker will demonstrate that this equivalence principle can be derived from simple energy considerations and therefore does not depend on the validity of the difficult theoretical fundamental which Berek gave for his theory. This energy basis allows also, for the first time, the demonstration of the following tacit assumption made in each diffraction theory of optical instruments, namely, that it is permissible to superpose diffraction optics on geometrical optics.

May I say a few words about truth in science? You all know geometrical optics and diffraction optics. At least, since the time of Fresnel and Fraunhofer, we have known that light has a finite wave-length, that therefore, geometrical optics is nothing more than an approximation, which is valid in all problems where the wave-length can be neglected. Notwithstanding, in all optical calculation work, we use purely geometrical optics and use diffraction optics only to explain the last minute details, for instance, the distribution of light in the image plane. And the results, our optical instruments, show that we are right. It would be very bad if we were forced to use diffraction optics, for the calculation of the diffraction effect produced by a single spherical refracting surface of finite aperture is a problem of extreme difficulty, which has not yet been solved correctly. Moreover, we know now from quantum optics, that diffraction optics itself is not the last word—not the finite truth.

In science, we always have to deal with approximations. We make assumptions, and then call for mathematics, our servant, to draw conclusions from these assumptions. Mathematics, itself, is true, but in applying it we have always to watch our step to see if the assumptions we have made are still valid in the problem we are dealing with, and if not, we have to make the necessary corrections.

In the following it is the aim of the speaker to show that we can compare the diffraction theory to Abbe's point of view in dealing with microscopic vision, whereas the equivalence principle of Berek, like geometrical optics, is the simpler theory which, however, is sufficiently valid only in those cases where a certain magnitude might be neglected.

The most difficult thing in dealing with a problem on which so much has been written is that the same words used by different writers do

---

* Communication No. 573 from the Kodak Research Laboratories.

† Presented by invitation at the annual meeting of the Optical Society of America held in Philadelphia, October 24–26, 1935.—EDITOR.

not have the same significance. I shall try, therefore, to give exact definitions of the terms used in this paper, and wish to ask you to direct your attention to these definitions, which may partially deviate from the meanings that are familiar to you.

Before going on to our subject, I shall call your attention to some theorems of geometrical optics, which may facilitate the understanding of our problems.

1. The duality principle

The author has published[1] a principle according to which, from any optical theorem, three others may be derived, which are simultaneously valid. If we restrict ourselves to the rays in the meridional plane of a rotationally symmetrical optical system, our theorem has the following significance: (Fig. 1). An arbitrary ray is defined by the height h at which it intersects the coordinate plane, and its angle σ with the optical axis. The corresponding image ray is given by h' and σ'. All theorems of geometrical optics can then be stated in the form of one or more equations between these magnitudes. The duality principle states now that from every correct equation comes a new correct equation if you replace h by $n \sin \sigma$, and $n \sin \sigma$ by $-h$; or h' by $n' \sin \sigma'$ and $n' \sin \sigma'$ by $-h'$; or make both substitutions together (n and n' are the refractive indices of object and image space, respectively).

This principle suggests the following classification of optical instruments:

1. *Telescopic systems:* Axial rays given by $\sigma = \sigma' = 0$. Object and image infinite.
2. *Human eye, photographic objectives:* Axial rays given by $\sigma = h' = 0$. Object infinite, image finite.
3. *Eyepieces:* Axial rays given by $h = \sigma' = 0$. Object finite, image infinite.
4. *Microscope, reproduction objective:* Axial rays given by $h = h' = 0$. Object and image finite.

This classification allows the results found for one type of instrument to be applied to all the others.

We see especially from this principle that if we designate the effective diameter of a telescope or photographic objective (twice the maximal height h of a parallel entering ray proceeding through the system) as the *object aperture*, A, of the telescope or photographic system, the corresponding value, i.e., half the *object aperture*, A, of an eyepiece or microscope system is given by $n \sin \sigma$, where σ is the maximal angle with the optical axis of a ray which proceeds through the entire system. Because the theorem is valid in diffraction optics too, it allows us to transfer the known laws of diffraction in telescopes to all other instruments.

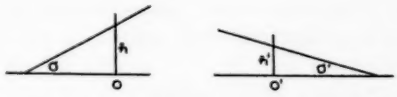


FIG. 1. The coordinates of a ray in object and image space.

2. The invariant of geometrical optics

Clausius demonstrated the following law (Fig. 2): We consider two surface elements, df_1 and df_2, which may be in different optical media (of n_1, n_2 indices). Let $d\Omega_1$, and $d\Omega_2$ be the solid angle which one surface subtends at an arbitrary point in the other, and u_1 and u_2 the angles which an arbitrary ray makes with the normals to the surfaces. We have the law of Clausius

$$n_1^2 df_1 d\Omega_1 \cos u_1 = n_2^2 df_2 d\Omega_2 \cos u_2. \qquad (1)$$

The expression on the left is proportional to the energy of the radiation that proceeds from the surface element. (It is equal to the energy if the intensity is unity.) Therefore, (1) is nothing else than an expression of the law of conservation of energy, if we neglect changes due to reflection and absorption. We have treated this theorem from the standpoint of energy, but it can be demonstrated as well by purely geometrical

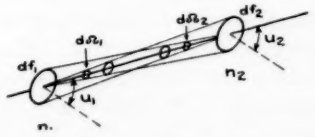


FIG. 2. Two surface elements which irradiate each other (not necessarily in the same space).

[1] Herzberger, J. O. S. A. **25**, 296 (1935).

optics.[2] The law stated above also holds true when the surface elements are conjugate.

3. The perfect optical system

We shall call an optical system perfect if it gives a sharp, flat, undistorted, point-for-point image of a flat object perpendicular to the axis of the system at the object point. Such an assumption stands in no contradiction to general optical laws, as does the often-used collinear object-image correspondence. It is possible to derive also from this assumption the well-known sine law of Abbe, which may be written in this form:

$$A = \beta' A', \tag{2}$$

where β' is the magnification of the system, and A, A' the object and image aperture, respectively. The duality principle shows how this formula changes when we consider another type of optical instrument.

These considerations allow us to restrict our investigations to the rays in the object space. If we know the light distribution in the object plane, the perfect instrument will do nothing more, from the standpoint of geometrical optics, than form a magnified image of this plane on the retina of the eye or eventually on a photographic plate. We now take our optical instrument, e.g., a microscope, and consider the object space. There are three planes in the object space where something of interest can happen: the plane of uniform illumination (not necessarily the image of the light source, for if we use Köhler illumination, it is the image of the first surface of the condenser), the plane of the object, and the entrance pupil of the optical system. If the object is a self-luminous body, plane of illumination and object plane are the same and only the diffraction at the margin of the entrance pupil need be investigated. If, however, the object is not self-luminous, the light from the illumination plane may be diffracted at the object, which we will consider for the moment as consisting of a great number of small diaphragms. It is now our task to consider under what circumstances we may neglect diffraction

[2] M. Herzberger, *Strahlenoptik* (Julius Springer, Berlin, 1931).

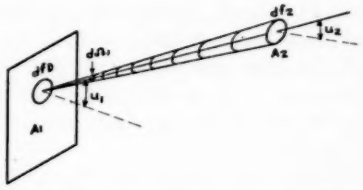


FIG. 3. The Airy disk seen through a small diaphragm on the object plane, if its normal forms an angle u_1 with the rays.

at the entrance pupil, and when we may neglect diffraction at the object.

Before we deal with a special case, we will remember the following: We will distinguish two kinds of diffraction phenomena. Fraunhofer investigated the diffraction which an infinite light source and a diaphragm or grating produce, if the plane of reference is at infinity. According to our duality principle, the investigation of the diffraction pattern in any plane, conjugated to the object point, and whether finite or infinite, can be dealt with in a corresponding manner. All these diffraction phenomena we call after Fraunhofer. If, on the other hand, we consider the diffraction phenomena at any other plane which is not in the neighborhood of a point conjugated to the object, we speak of Fresnel diffraction phenomena. Especially interesting are the phenomena at a plane conjugated to the grating.

We now consider the laws of diffraction regarding a simple special case. We will get a kind of differential law for diffraction problems, so that we may be able to get the general law by integration.

Given a luminous point A_1 (Fig. 3) and a small bundle of rays emerging from A_1, which is given by its solid angle $d\Omega_1$. We can, for instance, imagine this angle to be given by a circular diaphragm of the magnitude df_2, at A_2. The rays of our small bundle might have the angle u_1 against a given plane through A_1. Then we look at our object (with or without an optical instrument). At the plane through A_1 (or at a conjugate plane) we get a Fraunhofer diffraction phenomenon produced by light source A_1, and diaphragm df_2. We can regard df_2 as a small circle and we get the well-known result that, instead of our light point we see an Airy disk

which eventually is surrounded by light rings whose brightnesses decrease quickly. The intensity in the plane through A_1 can easily be given with the help of Bessel functions. We will restrict ourselves in considering only the inner circle, the so-called Airy disk. According to well-known formulae, the magnitude $df_d \cos u$ of its projection on a plane orthogonal to the rays is given by

$$df_D = C\pi^2\lambda^2/n^2 \cos u d\Omega. \tag{3}$$

The constant C is a physiological constant which, in most textbooks is written 1.22[2]. We can obtain a fairly close approximation if we set for our purposes, $C=1$. The well-known formulae for a circular aperture and an object plane orthogonal to the central ray can for instance be obtained by integrating (3). The sign $\backsimeq$ means approximately equal,

$$f_D \backsimeq \frac{\pi^2\lambda^2}{\int n^2 \cos u d\Omega} = \frac{\pi\lambda^2}{n^2 \sin^2 u} = \pi\left(\frac{\lambda}{A}\right)^2 = \pi r_D^2, \tag{4}$$

where r_d is the radius of the Airy disk. The Airy disk for all other diaphragms can be obtained in the same way. If there are several diaphragms (for instance, a grating), we can distinguish each Airy disk, and add the intensities.

Since now we have considered only a single luminous point, we will investigate a luminous little surface element, df. If we look at the object (or a conjugate plane), we see, instead of any point, according to our considerations an Airy disk, df_d, whose extension is given by (3). Now, we first consider that,

$$df \ll df_D. \tag{5}$$

In this case, all these Airy disks overlap in such a way that they are indistinguishable from a single Airy disk which has the mean intensity of all points of the light source. We are by no means able to distinguish such a light source from a point light source. We can treat this case as if the light source were a point. Particularly, we consider now two different points of the diaphragm. We can say that, through every point of the diaphragm comes only one ray. The rays coming to two points therefore have a certain path difference, and therefore a certain difference of phase. After the light has passed through the plane of the diaphragm, we get an interference phenomenon, as if the light were coming from a single point. The condition according to (5) and (3) is that

$$n^2 df \cos u d\Omega \ll \pi^2\lambda^2,$$

or

$$n^2 df \cos u d\Omega/\pi^2\lambda^2 \ll 1. \tag{6}$$

Comparing (6) with (1), we see that in (6) we have our optical invariant (1) and see that this quantity depends only on the light flux. We consider particularly two points, whose distance is smaller than or equal to r_d, the radius of the diffraction disk, then we cannot distinguish the points from a single point,—the two points are not resolved. We call your attention here to the well-known but sometimes forgotten fact that, we can speak of resolution only if we have two different points. A single point can always be made visible, if the intensity is not too small, but two points are not distinguishable if their distance is smaller than r_d.

We will now give a definition of what we call *coherent* light. We say that a light source gives *coherent* light if

$$0 < n^2 \int df \int d\Omega \cos u/4\pi^2\lambda^2 \leqq 1. \tag{7}$$

This magnitude is approximately equal to the light flux; in either case, the difference between the integral in (7) and the double integral over (1) can be neglected for the purposes of this paper.

We have got the first part of Berek's equivalence principle, a special case of which is given by T. Verdet in his lecture, and is to be found again in a paper of Lord Rayleigh.[3]

In general, we may find for a certain diaphragm and an extended light source, a number N given by the inequalities

$$N-1 < n^2 \int df \int d\Omega \cos u/4\pi^2\lambda^2 \leqq N. \tag{8}$$

In this case, we can regard our illumination plane as consisting of N radiating points (Fig. 4). N is equal to the product of aperture and field, divided by $(2\pi\lambda)^2$. N is an invariant of the

[3] Investigations in optics with special reference to the microscope.

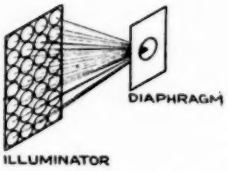


FIG. 4. Shows N number of Airy disks which fill our field, i.e., the number of details that can be distinguished, if field and entrance pupil are given.

optical system, and it is the number of Airy disks, with which you can fill out the illumination plane. If we think of a self-luminous object (object and illumination plane together), then N gives the number of details which can be made visible by the instrument. N is usually of the order of magnitude of 10^6. It is a number which is characteristic for an optical system. The progress in optics is the progress in increasing this number. That is the reason why, for instance, systems with a wide-angled field mostly have a small aperture, and systems with a large aperture cannot have a large field. This number N was, as far as I know, first suggested by E. Lihotzky in an unpublished lecture given at a meeting of the German Optical Society. It may be in close connection with the conception of M. von Laue of the degrees of freedom of an optical system.

Now, we shall apply our knowledge to the image formation of the microscope. We think first of our object as consisting of a series of small diaphragms and then consider each diaphragm and its effect. Two cases are of special interest: (1) If light source and diaphragms are so small that $N=1$, and (2), the case where N is a large number. It is obvious that the latter condition is chiefly fulfilled in microscopic practice, where you have a large illumination aperture, and it is obvious also that the first case approximately is fulfilled in the experiments of Abbe, in which he illuminated the object with a small beam of light.

First we consider the case of $N=1$. In this case we know that we can substitute for our light source a point, with the mean intensity of the points of the source. Our definition of coherent light was: Light such that it cannot be distinguished from light coming from a luminous point, and therefore able to interfere. From this definition you understand why it is possible to make interference experiments with the light coming from a star, which may be a body millions of times greater than the earth. You see at once that the points in the diaphragm have a path difference against each other, and send out coherent light. What you see, looking at the diaphragm with or without an instrument, is just the interference phenomenon caused by the light source in the plane of the diaphragm, a kind of Fresnel interference. This has more to do with the light source than with the diaphragm, and it is obvious that the same effect can be produced by different forms of the diaphragm, so that you get the results of Abbe's experiments.

I remember, for comparison, the method of Young, of making small particles visible by means of their diffraction pattern, which depends only on the *size* of the particles, not on their number or distribution. According to Babinet's theorem, a great number of small diaphragms would have the same effect as a great number of small particles.

The experiments of Abbe are of this kind. The object is illuminated with a small parallel bundle of rays, parallel to or at a finite angle with the axis, which means a point source at infinity. You may consider as an object a grating, single slit, double slit, or something similar. A Fresnel interference pattern is formed at the object, and the microscope does nothing but enlarge the image of this pattern. Of course, this statement would be valid only if all the diffracted light passed through the system. If the aperture of the system is not great enough, some of the light is excluded. That is the only effect the aperture of the instrument has in this type of experiment. If only the direct light and no diffracted light comes through the system, the image plane appears uniformly light, and shows no structure of the object. To show any structure, there must be at least one diffraction spectrum coming through the system. Abbe stated,[4] using a regular grating and some other objects, how great the aperture must be to allow two spectra to pass. You might keep in mind

[4] Lectures published by Lummer and Reiche, Braunschweig, 1910.

that in this case we do not have what we shall call an optical image of the object, but some kind of interference phenomenon. The image is not similar to the object, but is only related to it. From the image you can conclude that there is an object, you can investigate its qualities of symmetry, but there is no one-to-one correspondence between image and object. Unfortunately, the word *resolution* is used by Abbe and his successors for this case where two spectra pass through the system. We suggest using the term *interference resolution* to distinguish it from *optical resolution*. Unfortunately, for the microscope, the two figures are of the same order of magnitude and this has led to much confusion in the literature of the microscope.

Now we shall consider the other case, when N is a large number. Then, at each point of the diaphragm (or system of diaphragms) there arrive N primary rays coming from different points. If N is a large number, no phase will be distinguished, and the rays emerging from such a point in the diaphragm will be such that this point may be taken as a new light source. Of course, the radiation diagram for this point is not the same as that for a self-luminous point, but depends on the direct and diffracted light emerging from the point. However, that does not affect the conclusions which we may now draw. In this case, one or more bundles of rays emerge from every point in the object, and these rays meet in the image plane of the optical system, giving there an image just as if the object were self-luminous. We must therefore consider the diffraction effects due to the entrance pupil. We can now apply our previous conclusions regarding the microscope object as a light source and the entrance pupil of the system as a diaphragm, and we see that optical resolution is independent of the object and only a matter of the relative aperture of the optical system. A distance:

$$d \leqq \lambda/2A \tag{9}$$

cannot be resolved by an optical system.

We thus have got the second part of Berek's *equivalence principle*, which we now can describe as follows:

If $N=1$, the light source is equivalent to a single point with the mean intensity of all lighting points. If N is a great number, then we can regard the object as if it were self-luminous, that means, we can forget the special light source and consider the object as if each point were emanating coherent light. We only must take in consideration the radiation diagram of each object point which, of course, to a certain degree, depends upon the light source. We can get these results from the theory of Abbe if we integrate over the illuminating bundle.

That such a law must exist is easy to see. Nearly all our vision is vision of illuminated transparent or reflecting objects. Notwithstanding, we trust our vision to give us a true image of the object, unless the illumination is one-sided, irrespective of the light source.

Another example is the so-called Köhler illumination. To produce a uniform light source, you have only to illuminate the aperture of the condenser with a wide-angled cone of light and produce an optical image of this surface. This image shows none of the irregularities of the light source.

That there must be some limit where the special structure of the light source may be neglected can be seen by the examples mentioned, and by the image formed by the microscope when it is used, as usually, with high illumination aperture. The demonstration where this limit lies can be given in different ways. One demonstration which, however, is rather complicated can be given by the methods of Abbe's theory and other demonstrations can be given following an idea of Mandelstam. I sketch the demonstration briefly. If you have two planes, the plane of the light source and the plane of the diaphragm, you may ask for the light distribution in the plane of the diaphragm. If the radiation diagram for each point of the light source is known, you get the radiation diagram of a certain point in the diaphragm plane by solving an integral equation.[5] Especially if the diagram is a cosine radiation, you can ask under what conditions the radiation diagram at the plane is a cosine radiation also. In other words, under what conditions is the radiation reversible. It can be shown that this is the case only if

[5] Mandelstam, Heinrich Weber Festschrift.

$$4\pi^2\lambda^2/df \int d\Omega^2 \cos u \backsimeq 1/N$$

is negligible.

Now, in microscopic work, the assumptions of the second case are valid if we illuminate as is customary with a wide-angled cone of light, that is, with the condenser diaphragm wide open. In this case, the diffraction at the object, although always present, is negligible. Therefore, we get an image which accords with the laws of geometrical optics, except that every object point appears as an Airy disk.

Here, the investigations of Abbe about aperture and magnification are to be considered. The aperture of a system limits the distance which can be resolved. The magnification ought then to be chosen between two limits, the lower such that our eye, regarded as a camera with the pupil as diaphragm, is able to resolve this distance, and the upper such that the Airy disks are still seen as points. If we go further, we are likely to get a false impression of structure, arising from the rings surrounding the Airy disks.

If we compare this theory with the so-called Abbe theory, we find that there is the same relationship as that between geometrical optics and diffraction optics. The Abbe theory and diffraction optics give true results in all cases, but it is almost impossible to solve the more complicated problems with them because the mathematical difficulties are tremendous. It is a great help that we are able, in all problems where the wave-length may be omitted, to use geometrical optics; it is a great help that, in all problems where the light stream (flux) (energy for unit intensity) is great compared with λ^2, we can deal with the theory of the Airy disk, developed for the first time by Helmholtz, propagated by Gordon in England, and developed lately by M. Berek, who gave a very interesting and much disputed fundament to demonstrate his method. This paper will show that this theory may be based on a simple invariant of geometrical optics. If this were not true, the microscope would not be an instrument which could be used in so much scientific and industrial work. Only a few highly skilled mathematicians would be able to calculate the qualities of the object from the image they see, according to Abbe's theory, and also these would succeed only in a few simple examples.

In this short time, it is impossible to speak about all problems of microscope vision, but I wish to mention a few more points. First, the theory does not become much more complicated if we consider the actual optical system with all its image errors. The fundamental expression is an invariant of geometrical optics, so that we may consider it in any optical medium. That is the *fundament*, lacking until now, from which can be demonstrated that we get the diffraction in the image space by bending the wave which leaves the system according to geometrical optics, at the exit pupil of the system. This assumption, made in all theories of image formation, had never before been proved. And thus we can get in as simple a manner the Airy disk due to the actual emerging waves.

A second assumption was made when we dealt only with objects consisting of dark and light spots. That is, we assumed that the absorption in the object plane was either zero or infinite. Although that is seldom the case in real objects, it is for most purposes a sufficiently accurate approximation. If there is a continuous varying absorption, the influence on the image should be investigated, but I believe that the fundamental theorem holds true notwithstanding. That means that we get in this case too, a true image of the light distribution in the object plane, if our illuminating cones are chosen according to our formula.

We note further that, using wide-aperture illumination and reasonable magnification, we can get no more than a true and magnified image of the light distribution in the object plane of the microscope. If the object, as do many biological specimens, for instance, has very small variations in absorption and refractive index, it may not be seen under the usual conditions. But we cannot expect more because it is the same with macroscopic objects. A mirror or a window glass, if perfect and clean, is usually invisible. The methods of making these objects visible are by coloring and by special kinds of illumination which no longer give an exact image, but exaggerate some characteristics and differences. These methods, may I suggest, should be studied much more carefully with macroscopic

objects, which can be measured exactly. The results should be valuable in the study of corresponding microscopic objects.

Another problem: Some microscopic objects are not dielectric. This problem, too, should be investigated with macroscopic objects. One simple question not answered until now, as far as I know, is the following: Given a metallic bar of a certain known thickness, the problem is to determine this thickness by measuring an optical image, and investigate the influence of the illumination on the measurements.

We have taken from the wave optics only the fact that the wave-length is finite. Therefore, the theory cannot be applied to problems of polarization, or in brief, any problems that depend on the vector character of light. It can, however, be applied from the same reasons to all problems of wave mechanics, finally to all variation problems in which something is propagated with a finite wave-length. The correspondence of our laws to the laws of quantum mechanics would be more obvious if we had not dealt with the fiction of monochromatic light, but with light of a certain breadth of spectrum.

We have seen that, in the usual way of using the microscope, the simple Airy disk theory is sufficient. Nevertheless, the discovery of Abbe, that diffraction may be used on the object to produce interference spectra, is very valuable, and should receive more attention than it has heretofore. It is the first step on the way to making small distances *discernible*, which cannot be made *visible*. That is the direction in which Michelson went in his investigations, and which, recently, Kühne, Gerhard and Baeyer followed, using a specially modified microscope. Here may lie the direction of future progress. The usual method of microscopy has its limit: distances less than half a wave-length cannot be made visible. One way to reach beyond this limit in a small region is by dark-field illumination. Then, you do not see an image of the object, but an image of the diffraction pattern produced by the object. The other and much more far-reaching method is some form of interferometer, with a method of registering the intensity curves of the interference image.

This latter method has only one drawback: only a very small part of the incident light goes to form the image; and I believe that this is the reason that not very much about the object may be concluded from the interference pattern. If it were possible to build an apparatus which would bring all parts of the light proceeding from the object, perhaps successively, to an image, then, I venture to say, there would be a possibility of investigating all directions of the object successively. I have considered such an apparatus, but the difficulties of getting sufficient illumination are not easy to overcome, and the theory of the interpretation of the images is still rather complicated.

Now, in finishing, I should like to present a short survey of the history of the problems of microscopic vision, which is rather thrilling.

About 1822, Fraunhofer and Fresnel stated the theory of diffraction, which is the fundamental basis of all these theories. Fraunhofer at once considered the microscope, and remarked that objects smaller than one wave-length could not be resolved. These investigations were applied first to telescopic systems. After Herschel had made experiments on the real magnitudes of stars, Airy and Schwerd set forth, in 1835, the theory of diffraction in optical systems. Although Schwerd had discussed the problem in general, the microscopists heard nothing of these papers, written by astronomers. In the meantime, the construction of the microscope had been improved so that the fact of resolution was recognized. English opticians of that time discussed these phenomena, which depend on the aperture angle of the objective; and a paper written by I. Lister, in 1839, but published only recently by A. E. Conrady, shows that this clever scientist already knew that not the angle, but its sine, gives a direct measure of the resolving power. The discovery of the immersion objective, about 1850, by Hartnack in Germany, and Tolles in America, and the claim of the latter that immersion objectives had greater resolution, led to vehement discussion. (It is characteristic of the history of microscope vision that the discussions have been very vehement and spiced with personal attacks.) Tolles did not succeed in convincing his adversaries because he did not know the correct reason for his conclusions. But the theories of Abbe and Helmholtz, published in 1873, came to the cognizance of the

Royal Microscopical Society, and they discovered that both theories gave the limit of resolution as found before, and that both theories showed the importance of the numerical aperture, although they did not notice the difference in the point of view of the two theories. In his paper, Abbe sketched his theory and experiments, showing that the diffraction at the object is important, whereas Helmholtz treated the matter as if the object were self-luminous. Helmholtz acknowledged at once the importance of Abbe's theory, and wrote in a postscript that he might not have published his paper if he had not promised it for the jubilee volume of the *Poggendorff Annalen*. It is easy to understand with what enthusiasm this theory was received in England, if we read books about microscopy of this time, and see how impossible it was to explain well-known phenomena, such as resolution, for instance, where the authors made an attempt to distinguish between resolving power and a mystical penetrating power of the microscope.

Now comes the great development of microscopy, bound inseparably with the name of Abbe. I name as milestones the invention of the condenser, aplanatic correction of the objectives, the invention of the apochromat, fluorite systems, the homogeneous immersion objective, and the discovery of the rules of useful magnification.

However, some points of Abbe's theory were weak. From his experiments he had drawn, by logical argument, some conclusions which are not correct, and the discussion following is devoted to those problems. First, Abbe had constructed his condenser to produce extremely oblique illumination. But soon English amateurs, notably Nelson, used this condenser for wide-angled illumination. Abbe warned against this use, giving this argument: Extreme oblique illumination produces a certain interference spectrum, while axial illumination produces another interference spectrum. Both together give a kind of superposition of the spectra which is very dissimilar to the object. Therefore, it is not good to use wide-angled cones of illumination. The logical error may be seen if we consider an analogous conclusion. Suppose we make the simplest diffraction experiment possible, that is, consider the pattern formed by a pinhole and an infinite source. We see the usual pattern. Now, if we make another pinhole and fixate the light source through both, we see a diffraction effect which is very dissimilar to the light source. But if the hole grows to finite aperture. . . .

The other point is that Abbe's discovery of the nature of the microscopical image led him to exaggerate the differences between self-luminous and non-self-luminous bodies, and especially that the former could not show false structure.

The first attacks on Abbe's theory came from Altmann,[6] who took up the theory of Helmholtz, but believed that it also held true for the experiments of Abbe with small aperture, and that the diffraction and reflection at the object caused the effects seen by Abbe. Abbe answered this by giving a new series of experiments which demonstrated that diffraction at the object had no decisive effect.

Abbe had not published during his lifetime a mathematical demonstration of his theory, except the deductions of the diffraction effects of a few simple objects, given in his lectures at Jena. These lectures were first published in 1910 by *Lummer* and *Reiche*. Unfortunately, there is a mathematical error in this publication ($\cos \alpha$ instead of $(\cos \alpha)^{\frac{1}{2}}$), which was discovered by W. V. Ignatowsky.

Very few workers since this time have tried to calculate the interference patterns for special objects according to Abbe's theory. Abbe himself gave only the results for gratings. The first and most successful attempt in this direction was made in 1896 by Lord Rayleigh. He did not restrict himself to gratings, but investigated also a single point or slit, and a double point or slit. He investigated both self-luminous objects and the effects of illumination. His results show that the differences between self-luminous and non-self-luminous objects cannot be as important as Abbe thought, and "that the Abbe theory, while ingenious and capable of giving interesting results in certain directions, was not applicable to many of the problems which it is necessary to attack." The study of Lord Rayleigh's two papers should be obligatory for all who write on

[6] Altmann, Archiv. f. An. u. Phys., Leipzig, p. 111–84.

microscope vision. Gordon tried to apply the theory of Helmholtz to the problems of the microscope, but, as A. E. Conrady showed, he entirely ignored Abbe's experiments. The latter tried to develop Abbe's theory for finite aperture, but the results, which deal only with gratings, are quite complicated. A good step forward seems to the writer to be the address of Cl. Mercer, in 1896, before the American Microscopical Society on an experimental study of aperture as a factor in microscope vision, in which he shows that the diffraction experiments which Abbe had thought unique to the microscope could be performed with any other optical instrument. His experiments demonstrate, in agreement with Lord Rayleigh, that the use of a high aperture up to a certain limit is of advantage in microscopical work, and that in this way images can be produced of the same kind as the images of self-luminous bodies. We quote from his paper (p. 394).

> The more important results of our experimental study of aperture may now be summarized. It appears that diffracted rays leaving an object may be considered in the same category with other rays changed in direction by an object, and that the diffraction phenomena seen in a projected image are essentially the effect of changes in a light above the objective due to a function of the aperture, and not to changes below the objective due to diffraction of light in the plane of the object. It also appears, however, that diffraction in the plane of the object, does, under some circumstances, furnish light to certain parts of an aperture from which primary rays are absent, and thus enables the aperture more fully to determine the character of the projected image—resulting in a more nearly truthful image or, on the other hand, in false appearances; this is the gist of the Abbe phenomena of microscope vision. But it appears too, that such phenomena are not peculiar to microscope vision, notwithstanding Abbe's claim to the contrary, Moreover, with any positive lens similar and more brilliant results may be got by utilizing corresponding isolated pencils of primary rays instead of isolated pencils of diffracted rays. Still more trustworthy results may be got by using continuous apertures three-fourths (in diameter) full of primary rays instead of the isolated pencils of primary rays.

Very valuable also are the frequent discussions at the Royal Microscopical Society; it is only to be regretted that here, too, personal argument often took the place of mathematical or logical argument.

The work which Abbe had done was extended and improved by Russian authors, in particular Wolfke, Mandelstam and Ignatowsky. Mandelstam showed (theoretically and experimentally) that in certain circumstances a non-self-luminous body behaves like a self-luminous body. A new impulse came from a theorem of M. von Laue, who proved that this law is perfectly true if the body is illuminated from all sides.

Without comment I refer to a paper of H. Siedentopf (1912), in which he distinguished between diffraction laws for objects which have the same order of magnitude in all directions (surface elements), objects which are ultramicroscopic in one direction (line elements), and point objects. A later paper, dealing with the theorem of von Laue, considers the fact that no microscopical object can be illuminated from all sides as an argument that no microscopical object can be regarded as self-luminous. From our standpoint, this argument does not appear valid.

The influence of partial absorption on the diffraction disk was especially studied by *R.* Straubel, the influence of the image errors by Straubel, J. Picht, Strehl, and G. S. Steward. The paper of the latter is especially recommended because it gives an approximation according to the geometrical Seidel theory.

In the last fifteen years, M. Berek[7] has published a number of papers attacking the Abbe theory. His results are based on the so-called equivalence principle. He says that Abbe deals only with an uninteresting special case, and that in nearly all cases, the microscopic object acts like a self-luminous body. His theory is based on the interesting assumption that the coherence length is finite. He distinguished three kinds of coherence according to a formula which depends on a conception of Michelson. Almost all his results are given in this paper, only from the standpoint based on the laws of energy. The attacks which he directed against Abbe, to whom he does not give due credit, may have prevented his theory from being considered without prejudice. We recommend especially the study of his book.[7]

Much experimental work has been done in recent years. (We refer to the work of Gerhard, Cl. Münster, Lakemann and Groosmuller.) A good critical survey of the theory is to be found

[7] M. Berek, *Entwicklung und gegenwärtiger Stand der Lehre von der Abbildung im Mikroskop* (Marburger Sitzungsberichte, Marburg, 1926).

in L. C. Martin's[8] book. Martin also published a paper in 1930 describing interesting experiments on the influence of the illumination, and containing critical sidelights on both Abbe's theory and Berek's equivalence principle.

In closing, may I say that the purpose of this paper has been to present a survey of the different theories of microscope vision, and to show a simple method by which these phenomena may be explained from a new standpoint. The mathematical foundation of these theorems will be presented in a later paper.

I wish to thank the Bausch and Lomb Optical Company for the use of its library with its excellent collection of periodicals and books on the microscope.

[8] L. C. Martin, *Applied Optics* (I. Pitman & Sons, Ltd., London, 1932), Vol. 2.

A Method for Making and Testing Quarter-Wave Plates

RICHARD M. EMBERSON, *Physics Laboratory, University of Missouri, Columbia, Missouri*

(Received October 11, 1935)

A test for quarter-wave plates is described which requires only a light source and collimator, half-silvered mirror, Nicol prism, and full-silvered mirror. The test is sensitive and does not require a quarter-wave plate for comparison or other purposes. A mechanism is described for tilting mica sheets of approximately correct thickness so that they become effective quarter-wave plates.

RECENTLY the author has had occasion to prepare some quarter-wave plates. Of eight treatises on optics that were examined, only one[1] gave any directions for making and testing quarter-wave plates. Briefly, these directions suggest splitting mica into thin sheets and examining them between crossed Nicols. Those sheets which completely restore the light are to be designated as quarter-wave plates. A need was felt for a more sensitive method of testing. The author succeeded in devising a more sensitive test and then found it necessary to improve the technique of making quarter-wave plates. The ideas, while probably not new, are apparently not widely known.

A Test for Quarter-Wave Plates

Fig. 1 shows the arrangement for testing quarter-wave plates. The source, S, and filters, F, give light of the desired wave-length range. The light enters the aperture, A, which is located at the principal focus of the collimating lens, C. H, N, Q, and M are, respectively, a half-silvered mirror, a Nicol prism, the plate to be tested, and a full-silvered mirror. Of the collimated light that is incident upon H, half is transmitted and half is reflected. This reflected portion passes through N and Q, is reflected normally from M, again passes through Q and N, and then half is transmitted through H to the observer at E.

Let us suppose that the Nicol prism is oriented so that the light which it transmits is polarized in a plane perpendicular to the paper. If the quarter-wave plate is removed, the light is reflected from M and, since the plane of polarization has not been changed, this reflected light is transmitted by N and the observer sees a bright image of the aperture.

Now let us place a quarter-wave plate between N and M. As the quarter-wave plate is slowly rotated, the observer sees the image of the aperture suddenly fade and disappear. In fact, the image disappears for four positions of the quarter-wave plate during one complete rotation. These positions are those at which the fast direction of the quarter-wave plate makes an angle of 45° with the plane of polarization of the light transmitted by N. At any of these four positions the plane-polarized light is divided into two equal components, one parallel to and the other perpendicular to the fast direction of the quarter-wave plate. After passing through Q, the fast component is a quarter of a wave-length ahead of the slow component and the light is circularly polarized. The light is reflected at M as circularly polarized light, and the fast component is still parallel to the fast direction of Q. Thus, after again passing through Q, the fast component will be a half of a wave-length ahead of the slow component. The light will now be polarized in a plane perpendicular to that transmitted by N, so the light is completely quenched.

If the plate is too thick, the fast component of the transmitted light is $\lambda/4+\epsilon$ ahead of the slow component. The light will be reflected at M with the fast component still parallel to the fast direction of the plate. Thus, when the light has passed back through the plate, the fast component will be $\lambda/2+2\epsilon$ ahead of the slow component. The result, of course, is elliptically polarized light. When the observer turns the plate through one complete rotation, the aperture image passes through four brightness minima. These minima occur when the fast direction of the plate makes an angle of 45° with the plane of polarization of

[1] R. W. Wood, *Physical Optics* (Macmillan, 1934), third edition, p. 352.

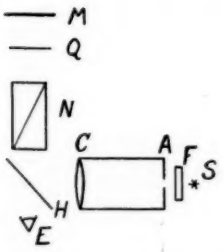


FIG. 1. A test for quarter-wave plates.

the light transmitted by N. At these four positions the elliptically polarized light is oriented with the major axis of the ellipse perpendicular to the plane of polarization which is transmitted by N. Thus, only the minor axis component is transmitted. It can be shown that the minor axis of the ellipse is proportional to $2\sqrt{2}\epsilon$. It should be pointed out that the brightness of the aperture image will vary with the square of the amplitude of the light.

Concerning the relative merit of this test, it is a well-known fact that a pair of Nicol prisms may be set to quench a beam of light with far greater accuracy than they may be set to give maximum transmission; also, the effect of an increase of intensity of the light beam will be more pronounced for the extinction setting of the Nicol prisms. The observations necessary for the quarter-wave plate test are essentially the same as those involved in crossing two Nicol prisms. In addition, since the test plate is made to check itself, i.e., since the light traverses the plate twice, the effect of any discrepancy of the test plate from a true quarter-wave plate will be doubled.

A METHOD FOR MAKING QUARTER-WAVE PLATES

In recent years several papers[2] have been published which describe methods of splitting mica. Benedicks' paper is the most detailed; he describes a general method for splitting mica under water and gives special instructions for splitting very thin sheets. These directions are not of much assistance in splitting mica sheets of a definite thickness because of the difficulty of controlling the thickness of the layer that is split off. The author spent considerable time attempting to make quarter-wave plates by simply splitting off layers of mica until a plate of the proper thickness was found. When tested with the previously described method, only about one in a hundred of these plates proved to be a close approximation to a quarter-wave plate, while with the method to be described below, over ten percent of the plates could be made into quarter-wave plates.

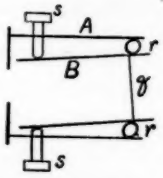


FIG. 2. Mechanism for tilting mica sheets.

Fig. 2 shows a simple arrangement for making quarter-wave plates. A sheet of mica, q, which has been split down a little too thin for a quarter-wave plate, is mounted in tube B. Tube B is placed in the larger tube, A; it is held centrally at one end by the ring, r, and the other end of B is held in position by set screws, ss. Tube B is tilted until a position is found for which the mica sheet acts as a quarter-wave plate for light moving parallel to the axis of tube A. It is easy to see that by tilting the plate slightly the lengths of the light paths of both the ordinary and extraordinary rays will be increased over their values for normal incidence; moreover, the light path of the extraordinary ray will be lengthened by a greater amount than the path of the ordinary ray. For some angle of tilt the path lengths will be such that the extraordinary ray will come out a quarter of a wave-length ahead of the ordinary ray, which is the necessary condition for a quarter-wave plate.

[2] Carl Benedicks, "Glimmer Spalten," Physik. Zeits. **32**, 561 (1931). See also Leif Tronstad, "An Improved Half-Shade System for the Detection of Elliptically Polarized Light," J. Sci. Inst. **11**, 146 (1934); John Strong, "On Splitting Mica," Rev. Sci. Inst. **6**, 243 (1935).

An Instrument for the Measurement of the Graininess of Photographic Materials *†

E. M. LOWRY, *Eastman Kodak Company*

(Received November 6, 1935)

The term "graininess" as applied in this paper refers to the subjective factor which has been well defined by Jones and Deisch as: "the sensation or impression of nonuniformity in a photographic deposit, produced upon the consciousness of the observer when such a deposit is viewed." The physical aspect of the problem, that is, the size of the ultimate silver grains, the size frequency and the number of grains per given area, has been defined in this laboratory as "granularity" and in this paper the term will be so used. Jones and Deisch used a method of measurement in which the observer determined the distance at which a magnified image of the deposit ceased to appear as nonhomogeneous. This is essentially the same as measuring the necessary magnification when the viewing distance is kept constant. It is the latter method which has been applied in the design of the instrument to be described. A number of methods for the measurement of graininess have been proposed at various times, among which have been the so-called physical methods, such as the measurement of Callier's "Q" factor or the scanning of the silver deposit and recording its nonhomogeneity by means of the response of a photoelectric cell or other light-sensitive device. These methods measure the granularity of the deposit rather than its graininess. In the last analysis the point of chief interest is the effect upon the consciousness of the observer, and so far as present knowledge is concerned a subjective method is the only one which gives promise of direct interpretation. Following the description of the apparatus and the defining equation some results showing the relation between field brightness and graininess are given for a range of brightnesses from 6 to 100 millilamberts. Data showing the effect of constant field brightness upon the graininess of several different emulsions are presented and a proposal for applying the results to the problem of graininess measurement is made. Results are also included on Q factor determinations as a measurement of graininess, together with a description of the apparatus used for this investigation.

INTRODUCTION

ALTHOUGH the size and distribution of the silver grains in a photographic deposit are of interest to the scientist, to the emulsion maker, and to others engaged in fundamental research on the properties of photographic materials, the practical photographer is concerned with an altogether different although undoubtedly closely related property of these materials. The appearance of the silver layer following moderate enlargement, such as occurs in the ordinary bromide enlargement, or on the motion picture screen, is the matter of primary interest to the photographer. With low magnification it is impossible to distinguish the individual grains which go to make up the emulsion, but there is apparent a lack of homogeneity in the deposit to which the term graininess has been applied.

The impression of graininess is the result not only of the grain size of the particles of which the deposit is composed, but also of the distribution and arrangement of these particles in clumps and of the further agglomeration of the clumps. Jones and Deisch[1] have well defined this subjective factor as, "the sensation or impression of nonuniformity in a photographic deposit, produced upon the consciousness of the observer when such a deposit is viewed." The appearance of nonhomogeneity at low magnification, as mentioned above, is due more to the presence of clumps of silver particles and to the massing together of these clumps than to the silver grains themselves. It is the granular structure of the silver layer which produces the shimmering and apparent boiling in a projected motion picture, or in making enlargements, prevents the photographer from making such enlargement more than a certain definite size. The physical aspect of the problem, that is, the size of the ultimate grains, the size frequency, and the number of grains per given area has been defined in this laboratory as "granularity" and in this paper the term will be so used.

Since the appearance of nonhomogeneity or

* Communication No. 572 from the Kodak Research Laboratories.

† Presented at the annual meeting of the Optical Society of America held in Philadelphia, October 24–26, 1935.—EDITOR.

[1] L. A. Jones and N. Deisch, "The Measurement of Graininess in Photographic Deposits," J. Frank. Inst. **190**, 657 (1920).

graininess depends to a great extent upon the resolving power of the retina, the method of measurement adopted by Jones and Deisch[1] was one in which the observer determined the distance at which a magnified image of the deposit ceased to appear as nonhomogeneous. Essentially, this is the same thing as measuring the magnification necessary when the viewing distance is kept constant. This latter method was suggested some years ago by Hardy and has been applied to the present investigation.

Following the appearance of the first papers by Jones and Deisch,[1] and later by Hardy and Jones,[2] on the subject of graininess and its measurement, a number of methods have been proposed. The method of Crabtree[3] consisted in projecting the image of a strip of exposed and developed film upon a screen and viewing the apparent boiling of the image. Then Conklin[4] developed the use of a specially constructed comparison microscope by means of which the graininess of the sample was matched with that of a grainy material instead of causing the graininess to disappear. Landau[5] also offered as a solution of the problem a method of comparison between the sample and a series of photographic enlargements of a grainy deposit.

Somewhat later Threadgold[6] proposed the application of Callier's so-called Q factor (ratio of specular to diffuse density) to the problem. Küster,[7] and then Küster and Eggert[8] followed the proposal of Threadgold for the use of the scatter coefficient with a slight modification of the latter's method in that they advocated the use of 100 log Q as the proper way to specify the graininess of any material. A further physical method of specifying graininess has been suggested which consists in scanning the silver deposit and recording its nonhomogeneity by means of the response of a photoelectric cell or other light-sensitive device. Such methods, of course, measure the granularity of the deposit rather than the graininess. It is quite likely that there is some relation between the physical method and the subjective one used in this laboratory, but as yet the exact relation is obscure. Any significant correlation of the two methods would require the collection of an enormous amount of data under very rigidly controlled conditions. Furthermore, the physical mode of attack upon the problem is subject to a considerable number of variables which may influence the results. In the last analysis the factor of chief interest is the effect upon the consciousness of the observer, and so far as present knowledge is concerned, a subjective method is the only one which gives promise of direct interpretation.

The Apparatus

The apparatus used in the work reported here is illustrated diagrammatically in Fig. 1. A small sphere, S, in which have been arranged a number of flash-light bulbs, B, provides a source of diffuse illumination. From 1 to 13 of the flash-light bulbs are used, depending upon the intensity level at which observations are to be

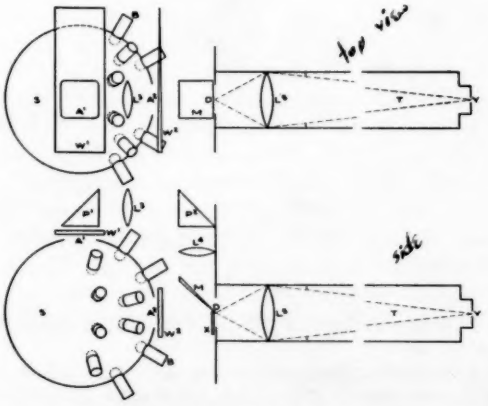


Fig. 1. Schematic diagram of graininess apparatus.

[2] A. C. Hardy and L. A. Jones, "Graininess in Motion Picture Negatives and Positives," Trans. Soc. Mot. Pict. Eng. No. 14, 107 (1922).

[3] J. I. Crabtree, "Graininess of Motion Picture Film," Trans. Soc. Mot. Pict. Eng. **11**, 77 (1927).

[4] O. E. Conklin, "Some Applications of the Comparison Microscope in the Film Industry," Trans. Soc. Mot. Pict. Eng. **16**, 159 (1931).

[5] R. Landau, "La Granulation," Techn. Cinématographique **1–2**, 8 (1930).

[6] S. D. Threadgold, "The Measurement of Graininess," Phot. J. **72** (N.S. 56), 348 (1932).

[7] A. Küster, "Über die Körnigkeit und das Auflösungsvermogen photographischer Schlichten," Phot. Korr. **70**, 17 (1934).

[8] J. Eggert and A. Küster, "Über die Körnigkeit photographischer Schichten," Kinotechnik **L16**, 127, 291 (1934).

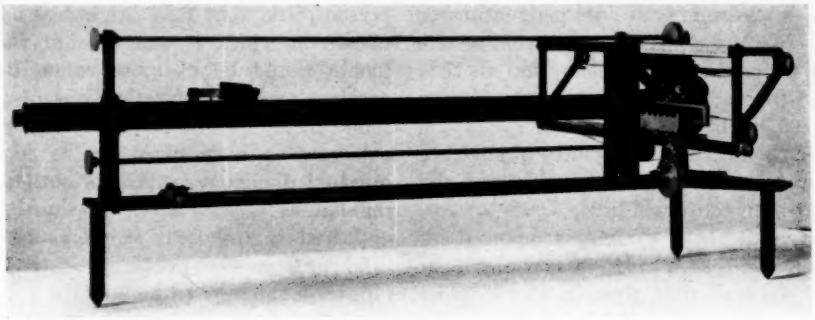


FIG. 2. Photograph of graininess apparatus.

made. Diffuse light from the sphere illuminates the photographic plate or film held in position at X. A textureless comparison field is formed by means of the prisms, P_1 and P_2, the lenses, L_3 and L_4, and the mirror, M. The mirror is of stellite which has been ground and polished to a sharp edge and this edge forms the dividing line in the photometric field. Intensity in the comparison field is controlled by the neutral wedge, W_1. W_2 serves as a means of setting the zero of the instrument. The lens L_5, in the telescope tube, T, is so mounted that it may be moved back and forth through a distance from the sample which is always less than its focal length. In this way the observer, Y, views a virtual magnified image of the deposit, X, the magnification depending upon the position of the lens with respect to the sample. The setting of the lens when the proper magnification has been found is indicated by a pointer and scale located on the telescope tube. A photograph of the apparatus is shown in Fig. 2.

One of the advantages of this form of apparatus is that the brightness of the field is sufficiently high that readings may be made without difficulty in a room lighted to moderate intensity, thus obviating the necessity of dark-adapting the observer's eye with the accompanying inconvenience and fatigue of working amid dark surroundings. The wedge, W_1, is so calibrated that the density of the deposit may be read simultaneously with the determination of graininess, thereby eliminating an additional operation. Refocusing is not required with each step of the sensitometric strip, the range of the instrument is greater, and the brightness higher than in the old instrument, thus making for increased accuracy of settings. Moreover, the angular dimensions of the field subtended at the eye by the lens remain practically constant throughout the entire range of the apparatus. The maximum magnification obtainable with the apparatus as now constructed is about 17.1 diameters. This, of course, means a range of 17 to 1, as opposed to a range of 8 to 1 in the old instrument. In other words, the graininess of a deposit which is 17 times as grainy as the least grainy deposit may be read.

In order that differences in the resolving power of the retinas of different individuals and differences in the same individual at different times might be eliminated, Jones and Deisch[1] referred the measurements of graininess to a half-tone screen having five hundred lines per inch. In this way instead of expressing the graininess of the film or plate in terms of the distance from eye to screen for mergence, the ratio of the distance for the half-tone screen to that of the film or plate was taken as a measure of graininess. The procedure is the same for the present equipment except that the ratio of the magnifications for mergence is used rather than the merging distance.

Defined on the above basis graininess is expressed by the equation

$$G=\frac{K}{n\times m_1/m_2},$$

where G is the graininess factor, K is an arbitrary

constant, n is the number of lines per centimeter (197 in this case) of the half-tone screen, m_1 the magnification for the film or plate, and m_2 that for the half-tone screen. The constant, K, was chosen as 100,000 by the previous investigators, so that even the finest-grained plates would have a graininess factor greater than unity, and the practice has been continued here.

The measurements of graininess reported by Jones and Deisch[1] were made at constant field brightness, since at that time it was thought that when reading a low density the intensity should be reduced, so that the total amount of light transmitted would be equal to that transmitted by a high density. It is easy to understand the reason for this conclusion, since the ordinary procedure in making photomicrographs, of a series of densities, is to alter the exposure so that about the same total amount of light is transmitted by all densities whether they be high or low. As a result of their method of measurement, the values of graininess were found to increase with increasing density. Data to be presented corroborate their findings.

Later, Hardy and Jones[2] altered the apparatus so that the intensity adjustment was in the comparison beam rather than in that incident on the material under examination. By this means the illumination of the sample to be examined was kept constant and the comparison beam adjusted for match. This method of illumination corresponds to ordinary conditions of viewing a deposit in which the field brightness as seen by the observer varies with the density. They found that with such an arrangement the maximum graininess appeared at densities of from 0.3 to 0.6 and decreased for both higher and lower densities. Experience bears out their findings, because, as usually viewed, graininess is most apparent in densities between 0.3 and 1.0, and, moreover, it approaches zero for very high and very low densities.

If the intensity of illumination on the negative is to remain constant it becomes necessary to choose, before starting the work, some value which may be kept constant throughout. Because of the limitations of their equipment, Jones and Hardy selected a value for the illumination of the sample of 7.6 footcandles. In the present investigation no such limitation was present, and that field brightness adopted was chosen at which it was thought the measurements would be of most value. Because the greatest need for investigations of this sort is found in the motion picture industry where the enlargement may often be as high as two hundred diameters, a field brightness of fifteen millilamberts was adopted, since that value appeared to represent very closely the screen brightness in the average theater. The magnification referred to here is the ratio between the size of the screen image and that of the transparency in the projector. In terms of the angle subtended at the eye of the observer, the average magnification is of the order of 4 or 5 to 1.

The Effect of Varying Field Brightness

The new apparatus was designed for working at considerably higher brightness levels than was possible with the old instrument, and for this reason it seemed advisable as a first step toward the study of graininess in general, that data be collected showing the relation between field brightness and graininess. The question has been investigated to a certain extent by Hardy and Jones,[2] but with their apparatus it was impossible to cover a brightness range sufficiently great to be conclusive. They found, however, that graininess increased quite rapidly with increasing field brightness.

For the present work, sensitometric strips were made on a fairly coarse-grained negative emulsion and readings of graininess were taken at field brightnesses of 6, 15, 50, and 100 ml. These values of brightness are for a clear field with no film in position. As demonstrated by inspection of Fig. 3, the effect is not only to shift the maximum of the graininess density curve toward the higher densities as the brightness level is increased, but also to broaden and straighten the curve itself. With still higher brightnesses the maximum graininess will no doubt move on toward higher and higher densities. However, there must be a certain density which will show no further increase in graininess no matter how high the brightness, since sufficiently high densities in the silver deposit will approach complete opacity and consequently zero graininess. It is unfortunate

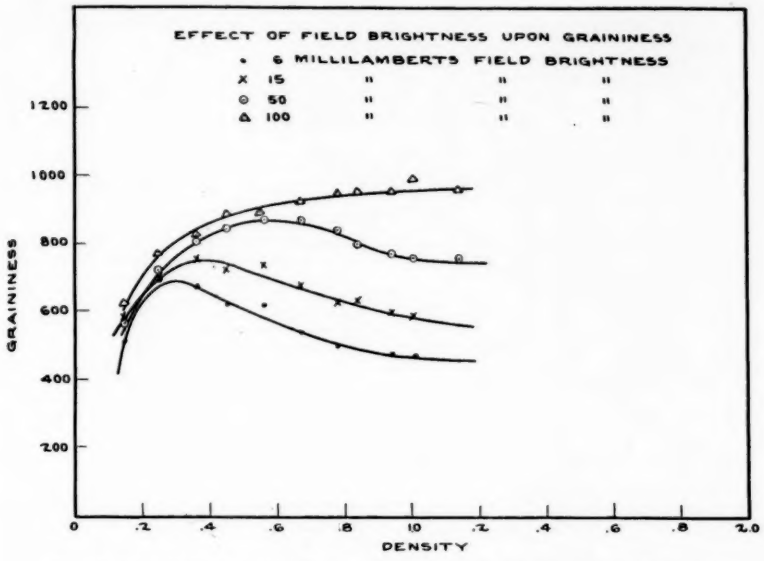


FIG. 3. Curves showing the effect of varying field brightness upon graininess.

that the sensitometric strips used in this work did not have densities of at least 2.0 but it is planned to repeat the experiment with fresh strips.

At this point it may be well to insert a brief description of the method of preparing the sensitometric strips. The reports of earlier workers in the field have shown that the presence of foreign material or abrasions on the emulsions is fatal to satisfactory measurements, and, accordingly, considerable care was exercised in the preparation of the samples. After exposure the film was brush-developed for a development time which yielded a gamma of unity. The developed strips were then quickly passed through a stop bath and fixed-out in clean, fresh, acid hypo. Following the fixing of the developed silver, the film was washed for thirty minutes in running water, swabbed off with cotton, rinsed in distilled water, and wiped dry with chamois. The drying process was completed in a dust-free drying cabinet. Careful manipulation, using this technique, provided test strips exceptionally free from defects of any sort.

THE EFFECT OF CONSTANT FIELD BRIGHTNESS

Previous data had indicated that if the field brightness were maintained constant, the value of graininess would increase with increasing density. To investigate this point further, sensitometric strips were prepared on four emulsions, namely, Hyper Press, Eastman 33, Process, and Slow Lantern. All of these emulsions were coated on glass. For the reasons stated earlier in this paper, the field brightness at which readings were made was fifteen millilamberts. Fig. 4 shows the results obtained, and, in addition, those for Super Speed Portrait Film and Ciné Positive Film. The previously reported results are substantiated by these data and a very interesting and perhaps useful fact is demonstrated. The plotted data for the various emulsions show the same straight-line relation between graininess and density, but each is represented by a line of different slope. If succeeding experiments bear out those so far secured, and there is no reason to suspect the contrary, it may be possible to classify emulsions according to the slope of this line. With the exception of the work of Jones and Deisch,[1] all measurements have been made on the basis of the maximum of the graininess density curve secured when the field brightness has been allowed to vary with the density. A graininess factor based on the slope of the graininess density curve at constant field brightness will

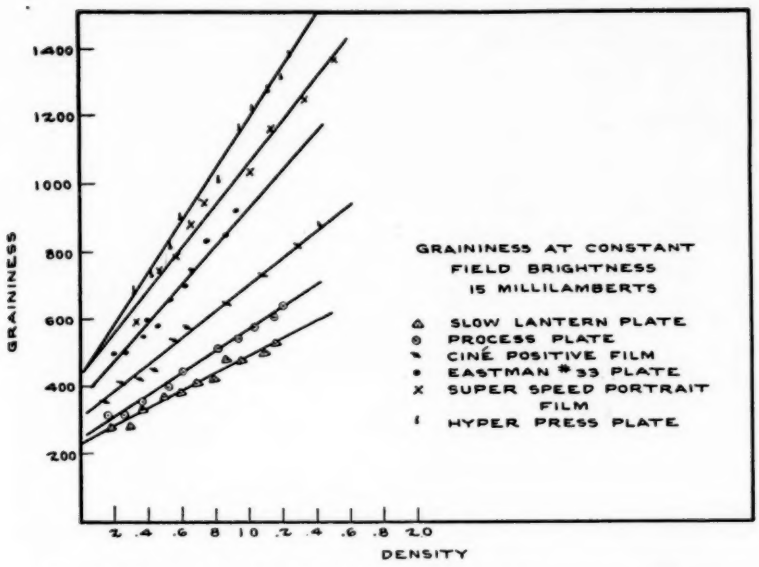


FIG. 4. Curves showing the effect of constant field brightness upon graininess.

have the advantage of permitting identification by readings made at perhaps three to five predetermined densities. If this can be done, the work of determining the graininess of various emulsions will be reduced approximately one-third to one-quarter of that required by the other method. It seems entirely logical to assume that the deposit exhibiting the maximum slope will also exhibit maximum graininess by the present method. The expression for graininess so secured, while it will not determine the maximum graininess for any density, will enable a definite factor to be applied to each emulsion. The straight-line relation is very similar to that reported by other workers using the so-called physical methods, and it may be that in this fact lies the connecting link between granularity and graininess.

Relation Between Callier's Quotient and Graininess

As mentioned earlier in this paper, Threadgold[6] proposed the use of Callier's quotient (Q) as a suitable and convenient measure for graininess. Later, Küster,[7] and then Küster and Eggert[8] reported on a slight modification of Threadgold's method in which they suggested that the quantitative expression should be one hundred times the logarithm of the Callier quotient, for a density of 0.5 measured in diffuse light. Jones and Deisch[1] had considered such a method some years previously, and since they found no general relationship between Q and graininess decided that such a method would not be satisfactory. In spite of additional recent evidence to the contrary, notably that of Narath,[9] who states that films differing widely in graininess yield nearly equal Q factors, Eggert and Küster have been quite insistent that the method is a suitable practical measurement of graininess.

It appeared worth while, therefore, to investigate the problem with data already at hand on the graininess of several emulsions as determined with the instrument just described, together with measurements made by the Q factor method. The apparatus used for the measurements of Q is one which was designed primarily for another purpose but which seemed admirably suited for this type of investigation. The instrument is shown in Figs. 5 and 6. Light

[9] A. Narath, "Über die Körnigkeit photographischer Schichten," Kinotechnik **16**, 255, 274, 287, 293 (1934).

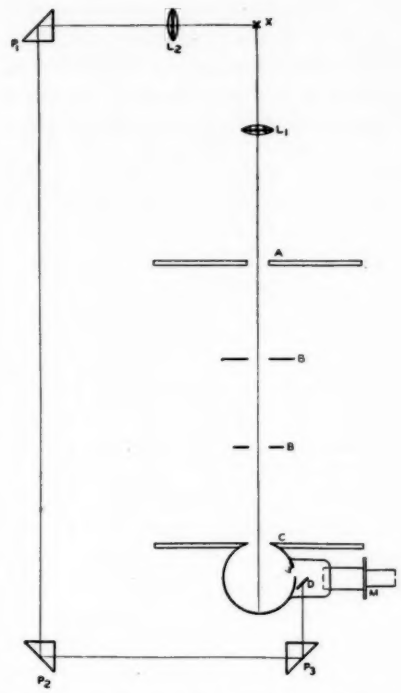


FIG. 5. Schematic diagram of the apparatus for the determination of Q factors.

FIG. 6. Photograph of apparatus for the determination of Q factors.

from the source, X, is collimated by the lens, L_1, and passes through the aperture, A, the diaphragm, B, and an aperture, C, into the integrating sphere, S. A comparison beam is formed by the lens, L_2, the prisms, $P_{1, 2, 3}$, and the magnesium oxide coated surface, D. The Marten's photometer, M, is used to examine the light from the sphere and from the surface at A. In making a determination of Q, the two beams of light entering the photometer are first adjusted for equality. Then the sample is placed over the aperture, A, and the specular density, $D_{||}$, measured. The aperture, A, is separated from C by a distance of twelve inches. This separation together with the collimation of the incident beam and the small angle subtended at the integrating sphere by the sample, insures the specularity of the density measured when the sample is at A. The sample is then placed over the aperture C in the integrating sphere and the diffuse density, $D_{\#}$ determined. This process was repeated for each step of the sensitometric strips previously used in the determination of graininess for each of seven emulsions. The values of $D_{||}$ were plotted against $D_{\#}$ for each type of emulsion and the value of $D_{||}$ for each at $D_{\#}$ equal to 0.5 were read from the curves.

In Table I are shown the values of the factor $K = 100 \log Q$ for each of the seven emulsions. Inspection of the values in the table shows that there is no significant relation between the Q quotient or the factor K determined from it and the graininess of the silver deposit. For

TABLE I. *Graininess values obtained by the visual and by the Q factor methods.*

TYPE OF EMULSION	GRAININESS	Q AT D 0.5	K
Supersensitive Panchromatic Film	810	1.80	25.5
Verichrome	765	1.86	26.9
Super Speed Portrait Film	605	1.83	26.5
Panatomic Film	600	1.72	23.5
Ciné-Kodak 8 Film	555	1.70	23.1
Ciné Positive Film	550	1.59	20.2
5C	448	1.17	6.9

example, Super Speed Portrait Film, which has a maximum graininess of 605, possesses a K value of 26.5, while Supersensitive Panchromatic Film, with a graininess of 810, has a K value of 25.5. Again, Ciné Positive and Ciné 8 emulsions, with equal graininess factors, show K values of 20.2 and 23.1. It is our conclusion from the evidence of these data that Callier's quotient is not a suitable means of determining the graininess of photographic materials.

The material presented in this paper is of a preliminary nature, having for its primary object a description of the new apparatus. For this reason the data are largely indicative of the applicability of the instrument to the investigation of the problem of graininess and of the correlation between the present and previously reported results. There are many factors related to the graininess of photographic materials which must be studied further and an extensive program of such studies is planned for the near future.

On a Method of Decreasing the Reflection from Nonmetallic Substances

JOHN STRONG, *California Instutute of Technology*
(Received September 25, 1935)

IN 1892, H. Dennis Taylor discovered a decrease in reflection associated with the tarnish which develops on flint glass surfaces. A thin plate, freshly polished, which reflected eleven percent from its two surfaces was observed to have its reflection decreased to five percent after tarnishing, resulting at the same time in a gain of six percent in transmission. Mr. Taylor found means to develop this tarnish chemically which, however, I do not believe has been described in the literature.

F. E. Wright[1] has experimented with various chemical treatments to produce this "tarnish." He suggests as an explanation of the effect that the surface of the glass is covered with small pits which are, compared with the wave-length of light, too small to scatter light. These pits result in a decrease of the refractivity of the material at the surface, giving a gradual transition from glass to air instead of an abrupt change in the index. This gradual transition leads to decrease reflectivity. Wright states that this phenomenon is similar to that exhibited by the zeolites. These minerals, as they lose their water of crystallization on the surface suffer a decrease in reflection coefficient.

The chemical treatments to produce tarnish have not had a wide-spread application to optical surfaces. This may be partly because of the danger involved of causing a permanent injury to the optical surface.

In this paper a physical method is described for producing the same effect as that produced by the chemical treatment. The method involves the coating of optical surfaces with an evaporated film of a nonmetallic substance. This film is removable. Some possible applications are described below.

It is well known that reflection from the convex surface of the collimator lens of a Littrow spectrograph gives a fogging of the photographic plate. This objectionable reflection may now be decreased several fold. The method may also find application in increasing the transmission of lenses and prism trains, as the light not reflected is saved in the transmitted beam.

Back-silvered plates and especially half-silvered plates frequently give secondary images due to reflection from the uncoated surface. In special cases, where it is important, these images may now be reduced in intensity.

Very fast lenses frequently give halation because of multiple reflections between the lens elements. This may also be reduced many fold in intensity.

Consider a transparent film deposited on an optical surface. The reflection will be materially changed when the index of refraction is different from that of the air and the underlying material. This can be illustrated by the application of Fresnel's reflection formula in the case of normal incidence for two nonmetallic substances of refractive indices n_2, of the film, and n_3, of the underlying substance:

$$R_{12}=\left(\frac{n_2-1}{n_2+1}\right)^2 \quad \text{and} \quad R_{23}=\left(\frac{n_3-n_2}{n_3+n_2}\right)^2.$$

R_{12} is the reflection at the film-air interface and R_{23} that at the other interface. The over-all reflection, R, will exhibit maxima and minima for varying film thickness. When $n_3=n_2^2$ the minimum value of R is zero and we would expect very low values of R to obtain over a fairly broad wave-length range. For ordinary substances the value of n_2 required to make R zero is about 1.25, a value less than that found for any known stable solid.

However, experiment shows that a substantial decrease in reflection is obtained by evaporating a film of fluorite onto glass ($n_3=1.52-1.53$) and magnesia ($n_3=1.75$). The reflection for these substances was tested before and after treatment. The reflection measurements were made using a 100-watt Mazda lamp as source and a Photronic cell as receiver. Assuming an index 1.43 for the fluorite film the computed decrease was 26 to 30 percent with the observed 26 percent. In

[1] Ordnance Department Document No. 2037, page 76.

this case R_{23} is very small (0.1 percent). For magnesia the expected decrease was 60 to 80 percent with the observed decrease 74 percent. Here R_{23} is one percent.

Another possibility of decreasing the reflection by the evaporation process involves the use of a film of decreasing index of refraction which starts out with the full index next to the glass. This has been realized by depositing a film of fluorite which is solid below and porous above. The grain of this porosity is small compared to a wave-length as evidenced by the fact that it does not scatter light. The porous film is obtained by evaporating at a pressure of 1/1000 to 1/10 mm of mercury. In one case such a film applied to a plate glass surface ($n=1.52$) gave a 54 percent decrease in reflection at normal incidence. In another case the decrease was 85 percent. At larger angles of incidence the reflection of the untreated glass grows considerably and the transmission is decreased correspondingly. Observations at about 80° incidence through a plate coated on both sides indicate that the transmission now is about double that of the untreated glass. This shows qualitatively that the transmission is increased at the same time that the reflection is decreased. This was also observed in the case of normal incidence.

Necrology

William Duane

WILLIAM DUANE, professor emeritus of biophysics at Harvard University, died at his home in Devon, Pennsylvania, on March 7, 1935.

Professor Duane was born in Philadelphia on February 17, 1872. He was graduated with the A.B. degree from the University of Pennsylvania in 1892, A.B. from Harvard in 1893, and A.M. from Harvard in 1895. As a Tyndall fellow from Harvard in 1895–97, he studied at Göttingen and Berlin, and received the Doctor of Philosophy degree from the University of Berlin in 1897. He was professor of physics in the University of Colorado from 1898 to 1907.

In 1907, Professor Duane with his family took up his residence in Paris and from 1907 to 1913 carried out, in the laboratory of Madame Curie at the Sorbonne, a number of precise and careful researches on the heat produced by radioactive disintegration, some chemical effects of the radioactive radiations and a method of registering individual particles. These papers were published in *Comptes rendus* and *Le radium*.

The physiological effects of the radioactive radiations created a great interest in their possible application to the treatment of disease, and in 1913 on the recommendation of the Harvard Cancer Commission and the Department of Physics, Professor Duane was called to Harvard as assistant professor of physics. In 1917, he was made professor of biophysics, a position he held until his retirement in September, 1934, when he became emeritus. From 1918, he was also research fellow in physics under the Cancer Commission.

Professor Duane was a member of many learned societies, including the American Physical Society, the Optical Society of America, Société de Physique, Philosophical Society, Radiological Society, American Academy of Arts and Sciences, National Academy of Sciences, Roentgen Ray Society, American Society for Cancer Research, and others.

For researches in radioactivity and x-rays he received the John Scott Medal in 1922, the Comstock prize of the National Academy in 1922, and the Leonard prize of the American Roentgen Ray Society in 1923. He was awarded honorary degrees of Doctor of Science by the University of Pennsylvania and the University of Colorado.

Professor Duane's researches were devoted largely to the fields of radioactivity and x-rays, and many of his researches had direct application to the treatment of malignant disease. However, he was primarily devoted to the field of atomic structure and the interaction of radiation and matter, some of his publications being: *The Measurement of X-Ray Wave Lengths, The Absorption and Emission of X-Rays by Various Materials as a Function of the Wave Lengths of the Rays, X-Ray Intensities as Related to the Atomic Number of the Anticathode, The Quantum Theory, Determination of Planck's Constant by Means of X-Rays, The Polarization of X-Rays.* He published more than ninety papers, either as sole author or as joint author with his students and associates.

Professor Duane was an excellent teacher and a great inspiration to a large number of advanced students with whom he came in contact. Many of the men and women who were privileged to work with him now occupy prominent positions in the leading educational institutions of America and abroad.

J. CRAMER HUDSON

Charles Elwood Mendenhall

CHARLES ELWOOD MENDENHALL, professor of physics at the University of Wisconsin, died on August 18, 1935. Born in Columbus, Ohio on August 1, 1872, the son of the distinguished physicist, Thomas Corwin Mendenhall, he was educated at Rose Polytechnic and Johns Hopkins, taking his doctorate under Rowland in 1898. After three years at Williams he went to Wisconsin, where he was rapidly promoted in rank, attaining a professorship in 1905. He was married in 1906 to Dr. Dorothy M. Reed, who, with two sons, survives him.

During the war Mendenhall was a major in the science and research division of the Signal Corps and later became chairman of the division of physical sciences of the National Research Council. He was a member of the National Academy, American Philosophical Society, and American Academy, a past vice president of the American Association for the Advancement of Science, and past president of the American Physical Society.

Mendenhall's research interests, while very broad, lay chiefly in the field of radiation. His first work of importance was a study of the black body radiation curve in the infrared. This led to other bolometric researches, including the design of one of the most sensitive galvanometers of that time. He collaborated with the late Augustus Trowbridge in a fundamental investigation of a phase of the ether-drift question connected with radiation. He originated the well-known V wedge black body, and was concerned for a number of years with measurements of the Planck and Stefan-Boltzmann radiation constants. His interests later turned to the photoelectric effect, where he and his students made marked advances by pushing the outgassing of the specimens to extreme limits. During the last few years he had also become interested in the problem of nuclear disintegration. His work was characterized by accuracy and thoroughness, and by conservatism in publication. His interest was not in the matter of how many

papers bore his name, but rather in how far each investigation could be pushed before publication became a real necessity.

Important as were Mendenhall's own researches, however, his greatest contribution to science was the advice and help and training given almost every hour of the working day over a period of half his lifetime to the group of young investigators working under him. The investigations in which he was in this way the guiding spirit have yielded results of no small importance in the field of modern physics.

L. R. INGERSOLL

Michael Idvorsky Pupin

MICHAEL IDVORSKY PUPIN, Ph.D., Sc.D., LL.D., Professor Emeritus of Electromechanics in Residence, in Columbia University, a life member of the Optical Society of America, died in New York City on March 12, 1935. His was a career of eminence in scientific investigation, in the application of science to invention and engineering, in teaching, in interpreting to the public the significance and value of scientific thought and work, and in public service of many kinds. His autobiography "From Immigrant to Inventor," published in 1923, a remarkable book that has had wide popular appeal, well reveals the man and tells the story of this son of Serbian peasants, prosperous and well esteemed in their community, who was born October 4, 1858 in the village of Idvor in the Austrian province of Banat, now a part of Jugoslavia. He arrived in this country in 1874, seized the first job that came his way, one on a Delaware farm, learned English and progressed from one job to another, learning from all his contacts, and glorying in the freedom and opportunity in this country. By 1879 he was able, scholastically and financially, to enter Columbia College where he won scholastic distinction in Greek, mathematics and physics. Immediately after graduation he was awarded a fellowship to enable him to go abroad to study physics, first at Cambridge where he worked with Routh, and two years later at Berlin under Helmholtz. With Helmholtz he worked on osmotic pressure and free energy, receiving the Ph.D. degree in 1889.

In the fall of 1889 he began his long career at Columbia University, attaining a full professorship in 1901 with the title of Professor of Electromechanics in the department of mechanics which was later at his request united with the department of physics. He retired from active teaching, but not from active university service in many ways, in 1929, becoming Professor Emeritus "in Residence," a position of very special honor in his university.

To refer to only three of his published papers, there was an important paper in 1894 on the tuning of electrical circuits, subsequently resulting in applications to radio tuning; a paper from his early work on x-rays, announcing in 1896 his discovery of secondary x-rays; in 1899 a fundamental paper on the mathematical theory of telephone lines. He made many inventions in the modern manner, by the application of the most erudite mathematical and experimental analysis and synthesis. Chief among these was his solution of how to add inductance to improve the transmission of long telephone lines, the development of the Pupin "loaded line." Until his last illness of a few weeks he was busily engaged in mathematical studies relating to electrical communication lines and in directing experimental investigations on these problems at the University in the Marcellus Hartley Laboratory, of which he was director.

Nothing was more marked in Pupin than his devotion to the American ideals of liberty, tolerance, and justice, and his pride in his adopted country. At the same time he was most loyal to his fellow citizens of Serbian origin and devoted to the welfare of his native Serbia. He was a leader in work with and for Serbians in this country, in many organizations and enterprises. Generous gifts went from him to Serbia for educational institutions and other social agencies.

He was the recipient of many honorary degrees, medals and decorations. The latest medal was the John Fritz Gold Medal, the highest American honor in engineering, awarded him in 1932 by the four largest societies of professional engineers.

While his work was mainly in other fields than that of optics or light, he was one of the first to understand fully and appreciate the electromagnetic theory of Maxwell, and light to the almost oriental imagination of Pupin was a living marvel to be studied. In an address—on the ideals and importance of the National Research Council—he charmed and stirred his audience with discourse in this vein: . . . "Just watch on some summer morning how the early rays of the sun arouse the slumbering rose from its blissful dreams, and remember that the rose responds because its body and soul are tuned to the melodies which the glorious sunlight is pouring into its enchanted ear. . . ."

A character of great strength, a personality of great attractiveness and of rare kindness, gave power and grace to his remarkable mentality. Only those who knew him can realize how justly the bronze tablet on the physics building of his university, now renamed the Pupin Physics Laboratories, inscribes him as "A true guide to the perplexed."

G. B. PEGRAM